Input file Fbh32142FL.seq; Output File 32142.trans
Sequence length 2660

```
                                                          M   A   A   T    4
CCTTTNTNRCCACGCGTCCGAGAGCGCCCCGCAGTCTTCGCGGAAAGCGTTCGGGGTAGGCG ATG GCT GCG ACG   12

     R   A   G   P   R   A   R   E   I   F   T   S   L   E   Y   G   P   V   P   E    24
    CGT GCA GGG CCC CGC GCC CGC GAG ATC TTC ACC TCG CTG GAG TAC GGA CCG GTG CCG GAG   72

     S   H   A   C   A   L   A   W   L   D   T   Q   D   R   C   L   G   H   Y   V    44
    AGC CAC GCA TGC GCA CTG GCC TGG CTG GAC ACC CAG GAC CGG TGC TTG GGC CAC TAT GTG  132

     N   G   K   W   L   K   P   E   H   R   N   S   V   P   C   Q   D   P   I   T    64
    AAT GGG AAG TGG TTA AAG CCT GAA CAC AGA AAT TCA GTG CCT TGC CAG GAT CCC ATC ACA  192

     G   E   N   L   A   S   C   L   Q   A   Q   A   E   D   V   A   A   A   V   E    84
    GGA GAG AAC TTG GCC AGT TGC CTG CAG GCA CAG GCC GAG GAT GTG GCT GCA GCC GTG GAG  252

     A   A   R   M   A   F   K   G   W   S   A   H   P   G   V   V   R   A   Q   H   104
    GCA GCC AGG ATG GCA TTT AAG GGC TGG AGT GCG CAC CCC GGC GTC GTC CGG GCC CAG CAC  312

     L   T   R   L   A   E   V   I   Q   K   H   Q   R   L   L   W   T   L   E   S   124
    CTG ACC AGG CTG GCC GAG GTG ATC CAG AAG CAC CAG CGG CTG CTG TGG ACC CTG GAA TCC  372

     L   V   T   G   R   A   V   R   E   V   R   D   G   D   V   Q   L   A   Q   Q   144
    CTG GTG ACT GGG CGG GCT GTT CGA GAG GTT CGA GAC GGG GAC GTC CAG CTG GCC CAG CAG  432

     L   L   H   Y   H   A   I   Q   A   S   T   Q   E   E   A   L   A   G   W   E   164
    CTG CTC CAC TAC CAT GCA ATC CAG GCA TCC ACC CAG GAG GAG GCA CTG GCA GGC TGG GAG  492

     P   M   G   V   I   G   L   I   L   P   P   T   F   S   F   L   E   M   M   W   184
    CCC ATG GGA GTA ATT GGC CTC ATC CTG CCA CCC ACA TTC TCC TTC CTT GAG ATG ATG TGG  552

     R   I   C   P   A   L   A   V   G   C   T   V   V   A   L   V   P   P   A   S   204
    AGG ATT TGC CCT GCC CTG GCT GTG GGC TGC ACC GTG GTG GCC CTC GTG CCC CCG GCC TCC  612

     P   A   P   L   L   L   A   Q   L   A   G   E   L   G   P   F   P   G   I   L   224
    CCG GCG CCC CTC CTC CTG GCC CAG CTG GCG GGG GAG CTG GGC CCC TTC CCG GGA ATC CTG  672

     N   V   V   S   G   P   A   S   L   V   P   I   L   A   S   Q   P   G   I   R   244
    AAT GTC GTC AGT GGC CCT GCG TCC CTG GTG CCC ATC CTG GCC TCC CAG CCT GGA ATC CGG  732

     K   V   A   F   C   G   A   P   E   E   G   R   A   L   R   R   S   L   A   G   264
    AAG GTG GCC TTC TGC GGA GCC CCG GAG GAA GGG CGT GCC CTT CGA CGG AGC CTG GCG GGA  792

     E   C   A   E   L   G   L   A   L   G   T   E   S   L   L   L   L   T   D   T   284
    GAG TGT GCG GAG CTG GGC CTG GCG CTG GGG ACG GAG TCG CTG CTG CTG CTG ACG GAC ACG  852
```

Fig. 1A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
    THEREFOR
Attorney/Agent: Kerri Pollard Schray
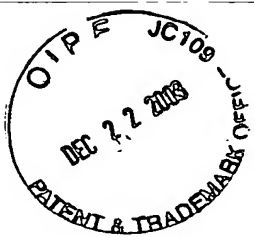Docket No.: MP100-079P1RCP2CN1M
Sheet 2 of 43
2/43

```
  A   D   V   D   S   A   V   E   G   V   V   D   A   A   W   S   D   R   G   P   304
GCG GAC GTA GAC TCG GCC GTG GAG GGT GTC GTG GAC GCC GCC TGG TCC GAC CGC GGC CCG  912

  G   G   L   R   L   L   I   Q   E   S   V   W   D   E   A   M   R   R   L   Q   324
GGT GGC CTC AGG CTC CTC ATC CAG GAG TCT GTG TGG GAT GAA GCC ATG AGA CGG CTG CAG  972

  E   R   M   G   R   L   R   S   G   R   G   L   D   G   A   V   D   M   G   A   344
GAG CGG ATG GGG CGG CTT CGG AGT GGC CGA GGG CTG GAT GGG GCC GTG GAC ATG GGG GCC 1032

  R   G   A   A   A   C   D   L   V   Q   R   F   V   R   E   A   Q   S   Q   G   364
CGG GGG GCT GCC GCA TGT GAC CTG GTC CAG CGC TTT GTG CGT GAG GCC CAG AGC CAG GGT 1092

  A   Q   V   F   Q   A   G   D   V   P   S   E   R   P   F   Y   P   P   T   L   384
GCA CAG GTG TTC CAG GCT GGT GAT GTG CCT TCG GAA CGC CCA TTC TAT CCC CCA ACC TTG 1152

  V   S   N   L   P   P   A   S   P   C   A   Q   V   E   V   P   W   P   V   V   404
GTC TCC AAC CTG CCC CCA GCC TCC CCA TGT GCC CAG GTG GAG GTG CCG TGG CCT GTG GTC 1212

  V   A   S   P   F   R   T   A   K   E   A   L   L   V   A   N   G   T   P   R   424
GTG GCC TCC CCC TTC CGC ACA GCC AAG GAG GCA CTG TTG GTG GCC AAC GGG ACG CCC CGC 1272

  G   G   S   A   S   V   W   S   E   R   L   G   Q   A   L   E   L   G   Y   G   444
GGG GGC AGC GCC AGT GTG TGG AGC GAG AGG CTG GGG CAG GCG CTG GAG CTG GGC TAT GGG 1332

  L   Q   V   G   T   V   W   I   N   A   H   G   L   R   D   P   S   V   P   T   464
CTC CAG GTG GGC ACT GTC TGG ATC AAC GCC CAC GGC CTC AGA GAC CCT TCG GTG CCC ACA 1392

  G   G   C   K   E   S   G   C   S   W   H   G   G   P   D   G   L   Y   E   Y   484
GGC GGC TGC AAG GAG AGT GGG TGT TCC TGG CAC GGG GGC CCA GAC GGG CTG TAT GAG TAT 1452

  L   R   P   S   G   T   P   A   R   L   S   C   L   S   K   N   L   N   Y   D   504
CTG CGG CCC TCA GGG ACC CCT GCC CGG CTG TCC TGC CTC TCC AAG AAC CTG AAC TAT GAC 1512

  T   F   G   L   A   V   P   S   T   L   P   A   G   P   E   I   G   P   S   P   524
ACC TTT GGC CTC GCT GTG CCC TCA ACC CTG CCG GCT GGG CCT GAA ATA GGG CCC AGC CCA 1572

  A   P   P   Y   G   L   F   V   G   G   R   F   Q   A   P   G   A   R   S   S   544
GCA CCC CCC TAT GGG CTC TTC GTT GGG GGC CGT TTC CAG GCT CCT GGG GCC CGA AGC TCC 1632

  R   P   I   R   D   S   S   G   N   L   H   G   Y   V   A   E   G   G   A   K   564
AGG CCC ATC CGG GAT TCG TCT GGC AAT CTC CAT GGC TAC GTG GCT GAG GGT GGA GCC AAG 1692

  D   I   R   G   A   V   E   A   A   H   Q   A   F   P   G   W   A   G   Q   S   584
GAC ATC CGA GGT GCT GTG GAG GCC GCT CAC CAG GCT TTC CCT GGC TGG GCG GGC CAG TCC 1752

  P   G   A   R   A   A   L   L   W   A   L   A   A   A   L   E   R   R   K   S   604
CCA GGA GCC CGG GCA GCC CTG CTG TGG GCC CTG GCG GCT GCA CTG GAG CGC CGG AAG TCT 1812
```

Fig. 1B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
       THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 3 of 43
3/43

```
    T   L   A   S   R   L   E   R   Q   G   A   E   L   K   A   A   E   A   E   V   624
   ACC CTG GCC TCA AGG CTG GAG AGG CAG GGA GCG GAG CTC AAG GCT GCG GAG GCG GAG GTG 1872

    E   L   S   A   R   R   L   R   A   W   G   A   R   V   Q   A   Q   G   H   T   644
   GAG CTG AGC GCA AGA CGA CTT CGG GCG TGG GGG GCC CGG GTG CAG GCC CAA GGC CAC ACC 1932

    L   Q   V   A   G   L   R   G   P   V   L   R   L   R   E   P   L   G   V   L   664
   CTG CAG GTA GCC GGG CTG AGA GGC CCT GTG CTG CGC CTG CGG GAG CCG CTG GGT GTG CTG 1992

    A   V   V   C   P   D   E   W   P   L   L   A   F   V   S   L   L   A   P   A   684
   GCT GTG GTG TGT CCG GAC GAG TGG CCC CTG CTT GCC TTC GTG TCC CTG CTG GCT CCC GCC 2052

    L   A   Y   G   N   T   V   V   M   V   P   S   A   A   C   P   L   L   A   L   704
   CTG GCC TAC GGC AAC ACT GTG GTC ATG GTG CCC AGT GCG GCC TGT CCT CTG CTG GCC CTG 2112

    E   V   C   Q   D   M   A   T   V   F   P   A   G   L   A   N   V   V   T   G   724
   GAG GTC TGC CAG GAC ATG GCC ACC GTG TTC CCA GCA GGC CTG GCC AAC GTG GTG ACA GGA 2172

    D   R   D   H   L   T   R   C   L   A   L   H   Q   D   V   Q   A   M   W   Y   744
   GAC CGG GAC CAT CTG ACC CGC TGC CTG GCC TTG CAC CAA GAC GTC CAG GCC ATG TGG TAT 2232

    F   G   S   A   Q   G   S   Q   F   V   E   W   A   S   A   G   N   L   K   P   764
   TTC GGA TCA GCC CAG GGT TCC CAG TTT GTC GAG TGG GCC TCG GCA GGA AAC CTC AAA CCG 2292

    V   W   A   S   R   G   C   P   R   A   W   D   Q   E   A   E   G   A   G   P   784
   GTG TGG GCG AGC AGG GGC TGC CCG CGG GCC TGG GAC CAG GAG GCC GAG GGG GCA GGC CCA 2352

    E   L   G   L   R   V   A   R   T   K   A   L   W   L   P   M   G   D   *       803
   GAG CTG GGG CTG CGA GTG GCG CGG ACC AAG GCC CTG TGG CTG CCT ATG GGG GAC TGA     2409
```

TGCCTGAGCGCCACCTACTGCATTTTGGACACCTCACACCAAGGGGAGATGCACCCCACAGACACCTGGGACTTTCCCC

TTCTGGTTCCTGTGTCTCCCAATAAACTCTCTGACCAACCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAARWARMAACTTC
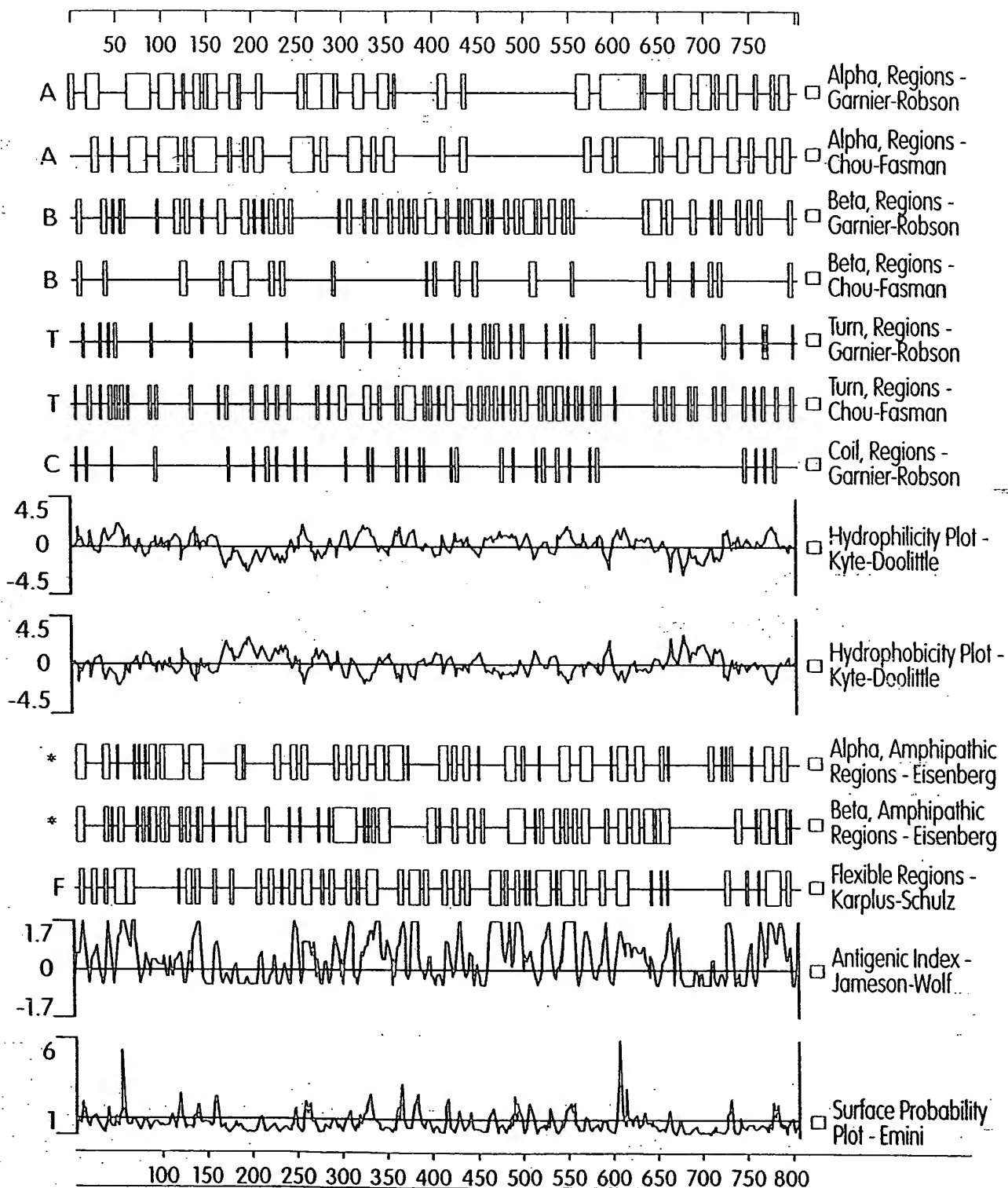
TGGCAGATATGAGGCTTTTTTCTTTTTTTTTT

Fig. 1C

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
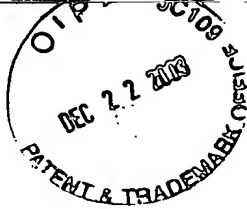Docket No.: MP100-079P1RCP2CN1M
Sheet 4 of 43
4/43

Fig. 2

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
159	175	ins-->out	0.1

>32142
MAATRAGPRAREIFTSLEYGPVPESHACALAWLDTQDRCLGHYVNGKWLKPEHRNSVPCQ
DPITGENLASCLQAQAEDVAAAVEAARMAFKGWSAHPGVVRAQHLTRLAEVIQKHQRLLW
TLESLVTGRAVREVRDGDVQLAQQLLHYHAIQASTQEEALAGWEPMGVIGLILPPTFSFL
EMMWRICPALAVGCTVVALVPPASPAPLLLAQLAGELGPFPGILNVVSGPASLVPILASQ
PGIRKVAFCGAPEEGRALRRSLAGECAELGLALGTESLLLLTDTADVDSAVEGVVDAAWS
DRGPGGLRLLIQESVWDEAMRRLQERMGRLRSGRGLDGAVDMGARGAAACDLVQRFVREA
QSQGAQVFQAGDVPSERPFYPPTLVSNLPPASPCAQVEVPWPVVVASPFRTAKEALLVAN
GTPRGGSASVWSERLGQALELGYGLQVGTVWINAHGLRDPSVPTGGCKESGCSWHGGPDG
LYEYLRPSGTPARLSCLSKNLNYDTFGLAVPSTLPAGPEIGPSPAPPYGLFVGGRFQAPG
ARSSRPIRDSSGNLHGYVAEGGAKDIRGAVEAAHQAFPGWAGQSPGARAALLWALAAALE
RRKSTLASRLERQGAELKAAEAEVELSARRLRAWGARVQAQGHTLQVAGLRGPVLRLREP
LGVLAVVCPDEWPLLAFVSLLAPALAYGNTVVMVPSAACPLLALEVCQDMATVFPAGLAN
VVTGDRDHLTRCLALHQDVQAMWYFGSAQGSQFVEWASAGNLKPVWASRGCPRAWDQEAE
GAGPELGLRVARTKALWLPMGD

Fig. 3

Protein Family / Domain Matches, HMMer Version 2

```
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:              /prod/ddm/seqanal/PFAM/pfam5.0/Pfam
Sequence file:         /prod/ddm/wspace/orfanal/oa-script.9519.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
  Query:  32142

Scores for sequence family classification (score includes all domains):
Model       Description                                      Score     E-value  N
--------    -----------                                      -----     -------  ---
aldedh      Aldehyde dehydrogenase family                    149.8     4.7e-41  1

Parsed for domains:
Model       Domain  seq-f,seq-t     hmm-f hmm-t       score  E-value
--------    ------  ----- -----     ----- -----       -----  -------
aldedh       1/1      47    494 ..      1   492 []     149.8  4.7e-41

Alignments of top-scoring domains:
aldedh: domain 1 of 1, from 47 to 494: score 149.8, E = 4.7e-41
                        *->ewvvdsasgktfevvNPankgevigrvpeataeDvdaAVkAAkeAfks
                           +w +++   + +++ +P + ge +++   +a+aeDv aAV AA+ Afk+
        32142     47     KWLWPEHRNSVPCQDPIT-GENLASCLQAQAEDVAAAVEAARMAFKG  92


                        GpwWakvpaseRariLrkladlieeredeLaaletlDlGKplaeAkgDte
                           W++ p    Ra+ L +la+ i+ ++ +L  le+l +G ++e+ + +
        32142     93     ----WSAHPGVVRAQHLTRLAEVIQKHQRLLWTLESLVTGRAVREVRDG-D  138


                        vgraideiryyagwarklmgerrvipslatdgdeelnytrrePlGVvgvI
                           v+ a + ++y a +a+          t+   e ++  +eP GV+g I
        32142    139     VQLAQQLLHYHAIQAS------------TQ---EEALAGWEPMGVIGLI  172


                        sPWNFPlllalwklapALAaGNTVVlKPSEqTPlt..alllaelieeaGa
                           P F +l ++w ++pALA G+TVV    + P+++ llla l  e G
        32142    172     LPPTFSFLEMMWRICPALAVGCTVV---ALVPPASpaPLLLAQLAGELG-  218


                        nnlPkGVvnvvpGfGaevGqaLlshpdidkisFTGSteVGklimeaAAak
                           +G +nvv G +a+ +   L+s+p+i+k++F G +e G+ + ++ A +
        32142    219     --PFPGILNVVSG-PASLVPILASQPGIRKVAFCGAPEEGRALRRSLAGE  265


                        nlkkVtLELGGKsPvIVfdDADLdkAverivfgaFgnaGQvCiApsRllv
                           +   L LG  s      d AD d Ave++v +a    G   ++ Rll+
        32142    266     -CAELGLALGTESLLLLTDTADVDSAVEGVVDAAWSDRG---PGGLRLLI  311


                        hesiydeFveklkervkklkliGdpldsdtniyGPlIseqqfdrvlsyIe
                           +es+ de + +l+er+ +l+  G +ld + + G+ +++ d v +++
        32142    312     QESVWDEAMRRLQERMGRLR-SGRGLDGAVDM-GAR-GAAACDLVQRFVR  358


                        dgkeeGAkvlcGGerdeskeylggGyyvqPTiftdVtpdMkImkEEIFGP
                           +++++GA+v + G ++ +    +    ++ PT++++++ p ++++ E+  P
        32142    359     EAQSQGAQVFQAGDVPSE---RP---FYPPTLVSNLPPASPCAQVEVPWP  402


                        VlpiikfkdldEAIelaNdteYGLAayvFTkdilarafrvakaleaGiVw
                           V++    f++  EA+  aN t+ G +a+v+++  l  a    +l++G+Vw
        32142    403     VVVASPFRTAKEALLVANGTPRGGSASVWSER-LGQALELGYGLQVGTVW  451


                        vNDvcvhaaepqlPFGGvHqSSGiGrehgGkygleeYteiKtVtirl<-*
                           +N    ++ +p++P GG K+ SG + ++ G++gl eY++     + rl
        32142    452     IN--AHGLRDPSVPTGGCKE-SGCSWHG-GPDGLYEYLRPSGTPARL    494
```

Fig. 4

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
    THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 7 of 43
7/43

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom 135 [Boxer ▼] [Showing match ▼] [Go!]	101	770	p99.2 (229) DHAL(10) DHAB(10) DHAM(7) // DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE	280
ProdomId	Start	End	Description	Score

View Prodom 135 [Boxer ▼] [Showing match ▼] [Go!]

>135 p99.2 (229) DHAL(10) DHAB(10) DHAM(7)   // DEHYDROGENASE OXIDOREDUCTASE
    ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE
    Length = 494

Score = 280 (103.6 bits), Expect = 7.8e-22, P = 7.8e-22
Identities = 87/289 (30%), Positives = 142/289 (49%)

```
Query:  216 ELGPFPGILNVVSG--PASLVPILASQPGIRKVAFCGAPEEGRALRRSXXXXXXXXXXXX 273
                E G  PG++NVV+G     A +   L S P I K++F G+ E G+A+ ++
Sbjct:  194 EAGLPPGVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMKAAAEKNLKPVTLE 253

Query:  274 XXXXX--XXXXDTADVDSAVEGVVDAAWSDRGP---GGLRLLIQESVWDEAMRRLQERMG 328
                     D  D+D AVE VV  A+ + G        R+ +QES++DE + +L ER+
Sbjct:  254 LGGKNPVIVFEDADDLDKAVESVVFGAFFNSGQVCTAASRIFVQESIYDEFVEKLVERVK 313

Query:  329 RL-RSGRG--LDGAVDMGAR-GAAACDLVQRFVREAQSQGAQVFQAGD---VPSERPFY- 380
               +L + G    LD   DMG       + +Q ++ EA+++GA++   G+       E   ++
Sbjct:  314 KLLKVGEDDPLDPDTDMGPLINEEQYEKIQSYIEEAKAEGAKLVCGGERRKAGDEGGYFI 373

Query:  381 PPTLVSNLPPASPCAQVEVPWPVVVASPFRT-AKEALLVANGTPRGGSASVWSERLGQAL 439
                PT++++++      Q E+ PV+   F+    EA+ +AN T G +A V++  + +A
Sbjct:  374 QPTILTDVTEDMRIMQEEIFGPVLPVIKFKDDLDEAIELANDTEYGLAAGVFTRDIERAQ 433

Query:  440 ELGYGLQVGTVWINA---HGLRDPSVPTGGCKESGCSWH-GGPDGLYEY 484
                 +   L+ GTVW+N   H  +    P GG K+SG    GG  GL EY
Sbjct:  434 RVAERLEAGTVWVNDNIYHVSAEAQAPFGGYKQSGIGGREGGKYGLEEY 482
```

Score = 262 (97.3 bits), Expect = 8.2e-20, P = 8.2e-20
Identities = 86/301 (28%), Positives = 140/301 (46%)

```
Query:  101 RAQHLTRLAEVIQKHQRLLWTLESLVTGRAVREVRDGDVQLAQQLLHYHA---------- 150
                RA+ L +LA+++++++  L  LE+L TG+ + E +  +V  A    L Y+A
Sbjct:   61 RARILRKLADLLEENKDELAALETLETGKPLAEAKVAEVARAVDYLRYYAGMAEKLMGEE 120

Query:  151 -IQASTQEE----ALAGWEPMGVIGLILPPTFSFLEMMWRICPALAVGVTXX---XXXXX 202
                I   S E      +     EP+GV+  I P F  + +W+I PALA G T
Sbjct:  121 TIPTSLSESPGSMSYTMREPLGVVAAITPWNFPLMMAVWKIAPALAAGNTVVLKPSEQTP 180

Query:  203 XXXXXXXXXXXXGELGPFPGILNVVSG--PASLVPILASQPGIRKVAFCGAPEEGRALRR 260
                            E G  PG++NVV+G     A +    L S P I K++F G+ E G+A+ +
Sbjct:  181 LTALLLAELIKEAEAGLPPGVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMK 240

Query:  261 SXXXXXXXXXXXXXXXX--XXXXDTADVDSAVEGVVDAAWSDPGP---GGLRLLIQESV 315
                +                      D  D+D AVE VV  A+ + G        R+ +QES+
Sbjct:  241 AAAEKNLKPVTLELGGKNPVIVFEDADDLDKAVESVVFGAFFNSGQVCTAASRIFVQESI 300

Query:  316 WDEAMRRLQERMGRL-RSGRG--LDGAVDMGAR-GAAACDLVQRFVREAQSQGAQVFQAG 371
                +DE + + +L ER+ +L + G    LD   DMG       + +Q ++ EA+++GA++   G
Sbjct:  301 YDEFVEKLVERVKKLLKVGEDDPLDPDTDMGPLINEEQYEKIQSYIEEAKAEGAKLVCGG 360

Query:  372 D 372
              +
Sbjct:  361 E 361
```

Fig. 5A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
   THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1 RCP2CN1M
Sheet 8 of 43

8/43

```
Score = 219 (82.2 bits), Expect = 4.9e-15, P = 4.9e-15
Identities = 75/236 (31%), Positives = 105/236 (44%)

Query:  550 SSGNLHGYVAEGGAKDIRGAVEAAHQAFPG--WAGQSP-GXXXXXXXXXXXXXXERRKSTL 606
            ++G +    V E   +D+  AVEAA +AF G  W   SP               E  K  L
Sbjct:   20 TNGEVIAQVPEATKEDVDKAVEAAREAFKGGEWGKTSPLSERARILRKLADLLEENKDEL 79

Query:  607 AS--RLERQGXXXXXXXXXXXXXXXXRRLRAW-GARVQAQGH-TLQVAGLRGP---VLRLRE 659
            A+   LE            LR + G  + G T+ +    P        +RE
Sbjct:   80 AALETLETGKPLAEAKVAEVARAVDYLRYYAGMAEKLMGEETIPTSLSESPGSMSYTMRE 139

Query:  660 PLGVLAVVCPDEWPLLAFVSLLAPALATGNTVVMVPSAACPLLAL---EVCQDMATVFPA 716
            PLGV+A + P  +PL+  V  +APALA GNTVV+ PS   PL AL    E+ ++     P
Sbjct:  140 PLGVVAAITPWNFPLMMAVWKIAPALAAGNTVVLKPSEQTPLTALLLAELIKEAEAGLPP 199

Query:  717 GLANVVTG-DRDHLTRCLALHQDVQAMWYFGSAQ-GSQFVEWASAGNLKPVWASRG 770
            G+ NVVTG   +   L  D D+ + + GS + G   ++ A+   NLKPV     G
Sbjct:  200 GVINVVTGFGGAEVGEALVSHPDIDKISFTGSTEVGKAIMKAAAEKNLKPVTLELG 255
```

Fig. 5B

Input file Fbh21481FL.seq; Output File 21481.trans
Sequence length 1379

TTTGGCCCTCGAGGCCAAGAATTCGGCACGAGGAGCAAGTGGCCTTAACACATGGATTTTCTTCCAAAAATGCAGACCC

ATTTTAATTAAGTTTGTAATTAACCACTGGGGAGGGCAGGCCCCCTGGATTCGGTCTGCTTTCGGAGACACTGTGAGTA

ACTTCCTATTTGTTGAACATTTGGGGATTAGCACGCCCACTGGGTGTTCAGCTTGGAGGCTTGCACAGAGCTGAGCTCC

CTGCAGCCTTGGGCCTCCCCCCTGCCCTGGGAGTCCTGATCAGCGTCTCTTTGCAAAGCCAATCCCCTTTTACTCCGTTG

```
                M   G   V   M   A   M   L   M   L   P   L   L   L   L   G   I    16
TCCCCCAGAACAAG ATG GGA GTC ATG GCC ATG CTG ATG CTC CCC CTG CTG CTG CTG GGA ATC   48

 S   G   L   L   F   I   Y   Q   E   V   S   R   L   W   S   K   S   A   V   Q    36
AGC GGC CTC CTC TTC ATT TAC CAA GAG GTG TCC AGG CTG TGG TCA AAG TCA GCT GTG CAG  108

 N   K   V   V   V   I   T   D   A   I   S   G   L   G   K   E   C   A   R   V    56
AAC AAA GTG GTG GTG ATC ACC GAT GCC ATC TCA GGA CTG GGC AAG GAG TGT GCT CGG GTG  168

 F   H   T   G   G   A   R   L   V   L   C   G   K   N   W   E   R   L   E   N    76
TTC CAC ACA GGT GGG GCA AGG CTG GTG CTG TGT GGA AAG AAC TGG GAG AGG CTA GAG AAC  228

 L   Y   D   A   L   I   S   V   A   D   P   S   K   T   F   T   P   K   L   V    96
CTA TAT GAT GCC TTG ATC AGC GTG GCT GAC CCC AGC AAG ACA TTC ACC CCA AAG CTG GTC  288

 L   L   D   L   S   D   I   S   C   V   P   D   V   A   K   E   V   L   D   C   116
CTG TTG GAC CTC TCA GAC ATC AGC TGT GTC CCA GAT GTG GCA AAA GAA GTC CTG GAT TGC  348

 Y   G   C   V   D   I   L   I   N   N   A   S   V   K   V   K   G   P   A   H   136
TAT GGC TGT GTG GAC ATC CTC ATC AAC AAT GCC AGT GTG AAG GTG AAG GGG CCT GCC CAT  408

 K   I   S   L   E   L   D   K   K   I   M   D   A   N   Y   F   G   P   I   T   156
AAG ATT TCT CTG GAG CTC GAC AAA AAG ATC ATG GAT GCC AAT TAC TTT GGC CCC ATC ACA  468

 L   T   K   A   L   L   P   N   M   I   S   R   R   T   G   Q   I   V   L   V   176
TTG ACG AAA GCC CTG CTT CCC AAC ATG ATC TCC CGG AGA ACA GGC CAA ATC GTG TTA GTG  528

 N   N   I   Q   G   K   F   G   I   P   F   R   T   T   Y   A   A   S   K   H   196
AAT AAT ATC CAA GGG AAG TTT GGA ATC CCG TTC CGT ACG ACT TAC GCT GCC TCC AAG CAC  588
```

Fig. 6A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
        THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 10 of 43

10/43

```
A   A   L   G   F   F   D   C   L   R   A   E   V   E   E   Y   D   V   V   I   216
GCA GCC CTG GGC TTC TTT GAC TGC CTC CGA GCC GAA GTG GAG GAA TAC GAT GTT GTC ATC 648

S   T   V   S   P   T   F   I   R   S   Y   H   V   Y   P   E   Q   G   N   W   236
AGC ACC GTG AGC CCG ACT TTC ATC CGG TCG TAC CAC GTG TAT CCA GAG CAA GGA AAC TGG 708

E   A   S   I   W   K   F   F   F   R   K   L   T   Y   G   V   H   P   V   E   256
GAA GCT TCC ATT TGG AAA TTC TTT TTC AGG AAG CTG ACC TAC GGC GTG CAC CCA GTA GAG 768

V   A   E   E   V   M   R   T   V   R   R   K   K   Q   E   V   F   M   A   N   276
GTG GCG GAG GAG GTG ATG CGC ACC GTG CGG AGG AAG AAG CAA GAG GTG TTT ATG GCC AAC 828

P   I   P   K   A   A   V   Y   V   R   T   F   F   P   E   F   F   F   A   V   296
CCC ATC CCC AAG GCC GCC GTG TAC GTC CGC ACC TTC TTC CCG GAG TTC TTT TTC GCC GTG 888

V   A   C   G   V   K   E   K   L   N   V   P   E   E   G   *               312
GTG GCC TGT GGG GTG AAG GAG AAG CTC AAT GTC CCG GAG GAG GGG TAA             936
```

CTGCAGGAGGCCAAATGGGCCACCCCTTGGAAATAAAGGTTTTTCTGGCAAAAAAAAAAAAAAAAAAAAAAAANTTTGCGGC
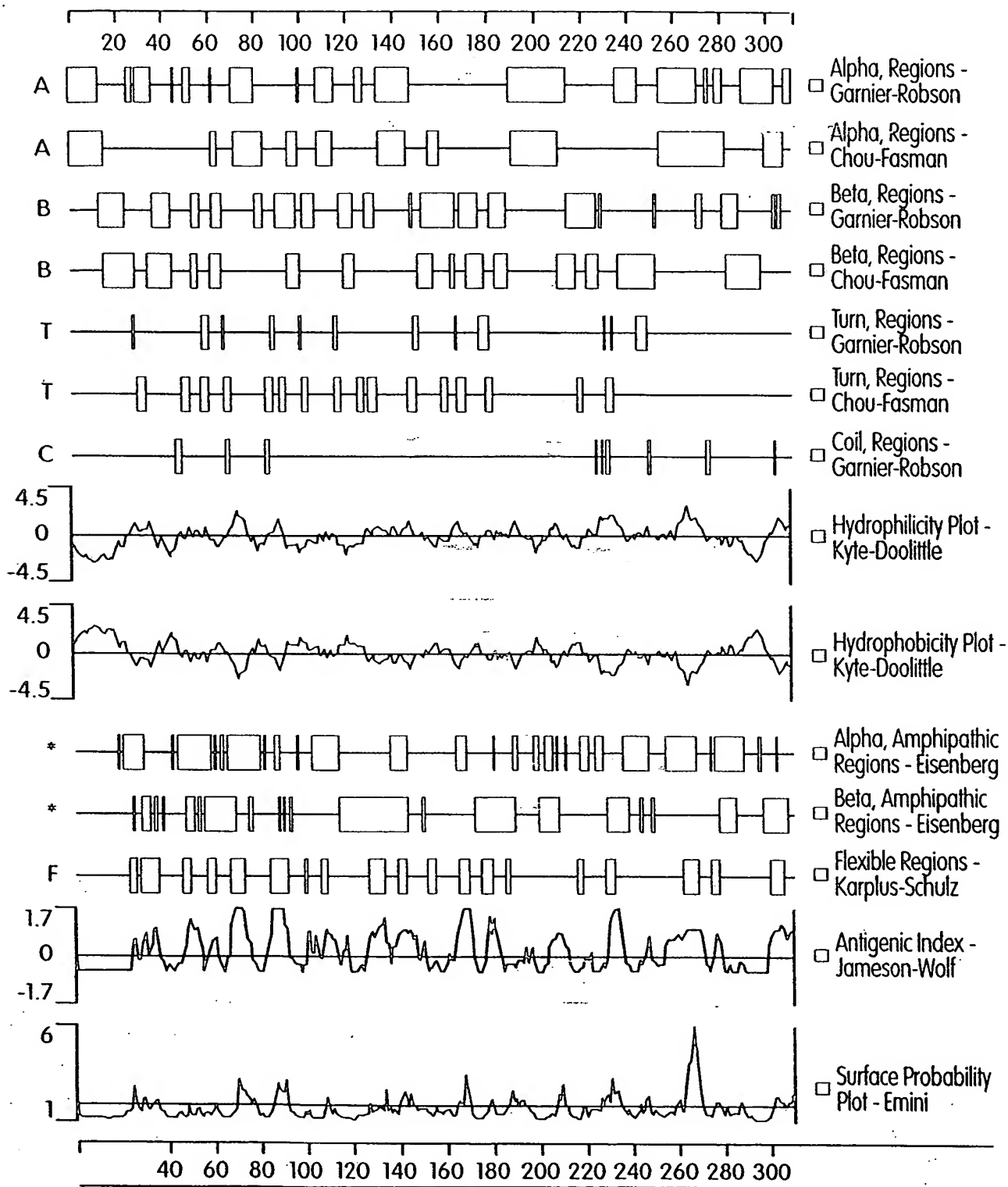
CGCAAGCTTATTCCCTTTAGGGAGGGGTTAATTTT

Fig. 6B

Fig. 7

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 12 of 43

12/43

Signal Peptide Predictions for 21481

Method	Predict	Score	Mat@
SignalP (eukaryote)	YES		19

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
7	23	ins-->out	4.5

>21481
MGVMAMLMLPLLLLGISGLLFIYQEVSRLWSKSAVQNKVVVITDAISGLGKECARVFHTG
GARLVLCGKNWERLENLYDALISVADPSKTFTPKLVLLDLSDISCVPDVAKEVLDCYGCV
DILINNASVKVKGPAHKISLELDKKIMDANYFGPITLTKALLPNMISRRTGQIVLVNNIQ
GKFGIPFRTTYAASKHAALGFFDCLRAEVEEYDVVISTVSPTFIRSYHVYPEQGNWEASI
WKFFFRKLTYGVHPVEVAEEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFFAVVACG
VKEKLNVPEEG

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
265	283	ins-->out	0.2

>21481 mature
LLFIYQEVSRLWSKSAVQNKVVVITDAISGLGKECARVFHTGGARLVLCGKNWERLENLY
DALISVADPSKTFTPKLVLLDLSDISCVPDVAKEVLDCYGCVDILINNASVKVKGPAHKI
SLELDKKIMDANYFGPITLTKALLPNMISRRTGQIVLVNNIQGKFGIPFRTTYAASKHAA
LGFFDCLRAEVEEYDVVISTVSPTFIRSYHVYPEQGNWEASIWKFFFRKLTYGVHPVEVA
EEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFFAVVACGVKEKLNVPEEG

Fig. 8

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 13 of 43
13/43

Protein Family / Domain Matches, HMMer Version 2
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- -
HMM file: /prod/ddm/seqanal/PFAM/pfam5.0/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.9650.seq
- -
 Query: 21481

Scores for sequence family classification (score includes all domains):
Model   Description   Score   E-value N
--------   -----------   -----   ------- ---
adh_short   short chain dehydrogenase   120.0   4.5e-32 1
A2M   Alpha-2-macroglobulin family   0.5   7.1 1

Parsed for domains:
Model   Domain   seq-f seq-t   hmm-f hmm-t   score E-value
--------   ------   ----- -----   ----- -----   ----- -------
adh_short   1/1   38   227 ..   1   203 [] 120.0 4.5e-32
A2M   1/1   278   291 ..   1   14 [.   0.5   7.1

Alignments of top-scoring domains:
adh_short: domain 1 of 1, from 38 to 227: score 120.0, E = 4.5e-32
                  *->KvaLvTGassGIGlaiAkrLakeGakVvvadrneeklekGavakelk
                     Kv+++T a sG+G+++A+ +++ Ga++v+++ n e+le+   ++l
       21481   38 KVVVITDAISGLGKECARVFHTGGARLVLCGKNWERLEN--LYDALI 82

                     elGgnd..kdralaiqlDvtdeesv.aaveqaverlGrlDvLVNNAGgii
                     +++++   + 1D++d + v+++++++++ +G +D+L+NNA   +
       21481   83 SV-ADPskTFTPKLVLLDLSDISCVpDVAKEVLDCYGCVDILINNAS--V 129

                     llrpgpfaelsrtmeedwdrvidvNltgvflltravlplmamkkrggGrI
                     gp++++s   +e+ ++++d N++g++ lt+a+lp   m+ r+ G I
       21481   130 -KVKGPAHKIS---LELDKKIMDANYFGPITLTKALLP--NMISRRTGQI 173

                     vNiSSvaGrkegglvgvpggsaYsASKaAvigltrsLAlElaphgIrVna
                     v + + G   + g p+++ Y+ASK+A g+ ++L+ E+ ++ + ++
       21481   174 VLVNNIQG-----KFGIPFRTTYAASKHAALGFFDCLRAEVEEYDVVIST 218

                     VAPGgvdTd<-*
                     v+P +++
       21481   219 VSPTFIRSY   227

A2M: domain 1 of 1, from 278 to 291: score 0.5, E = 7.1
                  ->idedditiRSyFPE<-
                     i+   + +R++FPE
       21481   278   IPKAAVYVRTFFPE   291

Fig. 9

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
   THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 14 of 43

14/43

ProDom Matches

ProdomId	Start	End	Description	Score
View Prodom 11 [Boxer ▼] [Showing match ▼] [Go!]	99	219	p99.2 (1078) ADH(34) GALE(20) FABG(13)// OXIDOREDUCTASE PROTEIN DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE	113
ProdomId	Start	End	Description	Score

View Prodom 11 [Boxer ▼] [Showing match ▼] [Go!]

```
>11 p99.2 (1078) ADH(34) GALE(20) FABG(13)  // OXIDOREDUCTASE PROTEIN
    DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE
    Length = 269

 Score = 113 (44.8 bits), Expect = 0.00016, P = 0.00016
 Identities = 41/138 (29%), Positives = 63/138 (45%)

Query:    99 DLSDIS-CVPDVAKEVLDCYGCVDILINNASVKV-KGPAHKISLELD-----KKIMDANY 151
              D+ D+     V V +E      +G +D+L+NNA V   K  A  ++ E       +++++ N
Sbjct:    87 DVEDVEKLVETVVEEFSGIHGKIDVLVNNAGVMAPKAVAESMTEETSDDEEWEEVIEVNV 146

Query:   152 FGPITLTKALLPNMIS--------RRTGQIVLVNNIQGK-FGIP-FRTTYAASKHAALGF 201
              G   LT+A LP M          R  G IV V ++ G   G P +   Y+ASK A   F
Sbjct:   147 TGTFNLTQAALPAMKKFSDAAAKKRFVGTIVNVASVAGSTMGSPGSQAAYSASKAAVESF 206

Query:   202 FDCLRAEVEEYDVVISTV 219
                 L  E+  Y   ++ V
Sbjct:   207 TKSLAMELSPYSASVAMV 224
```

Fig. 10

Input file Fbh25964Fl.seq; Output File 25964.trans
Sequence length 1725

GAGAAGGAGGAGCCAGCGGAAGGACGGTGTGCGGGCCGGCCAGCCCTGGACGAAAGAAGAGGGCCCCTCCAGGCCAGTC

TGGGCACCCTGGGATAGCGGCTGCAGCCATCAGCAGGGGCAGACGGCAGGTGGCCTGGTTGCTGCAGCTCCCAGGATCA

GCTCTGCCCTCCCCGCAAACGCCAGCCTCGTCACCGCTCCAGGGCACCTCCAGCAGTAACAGGTGGTTGCAGCAGGTGG

```
                                                        M   A   D   S   A   Q   A   Q   K       9
CAGCCAGCCCCTGGATGAGCCAAGGTCTCTTCCCCAGCCAGGC ATG GCC GAC TCT GCA CAG GCC CAG AAG     27

    L   V   Y   L   V   T   G   G   C   G   F   L   G   E   H   V   V   R   M   L      29
   CTG GTG TAC CTG GTC ACA GGG GGC TGT GGC TTC CTG GGA GAG CAC GTG GTG CGA ATG CTG    87

    L   Q   R   E   P   R   L   G   E   L   R   V   F   D   Q   H   L   G   P   W      49
   CTG CAG CGG GAG CCC CGG CTC GGG GAG CTG CGG GTC TTT GAC CAA CAC CTG GGT CCC TGG   147

    L   E   E   L   K   T   G   P   V   R   V   T   A   I   Q   G   D   V   T   Q      69
   CTG GAG GAG CTG AAG ACA GGG CCT GTG AGG GTG ACT GCC ATC CAG GGG GAC GTG ACC CAG   207

    A   H   E   V   A   A   A   V   A   G   A   H   V   V   I   H   T   A   G   L      89
   GCC CAT GAG GTG GCA GCA GCT GTG GCC GGA GCC CAT GTG GTC ATC CAC ACG GCT GGG CTG   267

    V   D   V   F   G   R   A   S   P   K   T   I   H   E   V   N   V   Q   G   T     109
   GTA GAC GTG TTT GGC AGG GCC AGT CCC AAG ACC ATC CAT GAG GTC AAC GTG CAG GGT ACC   327

    R   N   V   I   E   A   C   V   Q   T   G   T   R   F   L   V   Y   T   S   S     129
   CGG AAC GTG ATC GAG GCT TGT GTG CAG ACC GGA ACA CGG TTC CTG GTC TAC ACC AGC AGC   387

    M   E   V   V   G   P   N   T   K   G   H   P   F   Y   R   G   N   E   D   T     149
   ATG GAA GTT GTG GGG CCT AAC ACC AAA GGT CAC CCC TTC TAC AGG GGC AAC GAA GAC ACC   447

    P   Y   E   A   V   H   R   H   P   Y   P   C   S   K   A   L   A   E   W   L     169
   CCA TAC GAA GCA GTG CAC AGG CAC CCC TAT CCT TGC AGC AAG GCC CTG GCC GAG TGG CTG   507

    V   L   E   A   N   G   R   K   V   R   G   G   L   P   L   V   T   C   A   L     189
   GTC CTG GAG GCC AAC GGG AGG AAG GTC CGT GGG GGG CTG CCC CTG GTG ACG TGT GCC CTT   567

    R   P   T   G   I   Y   G   E   G   H   Q   I   M   R   D   F   Y   R   Q   G     209
   CGT CCC ACG GGC ATC TAC GGT GAA GGC CAC CAG ATC ATG AGG GAC TTC TAC CGC CAG GGC   627

    L   R   L   G   G   W   L   F   R   A   I   P   A   S   V   E   H   G   R   V     229
   CTG CGC CTG GGA GGT TGG CTC TTC CGG GCC ATC CCG GCC TCT GTG GAG CAT GGC CGG GTC   687
```
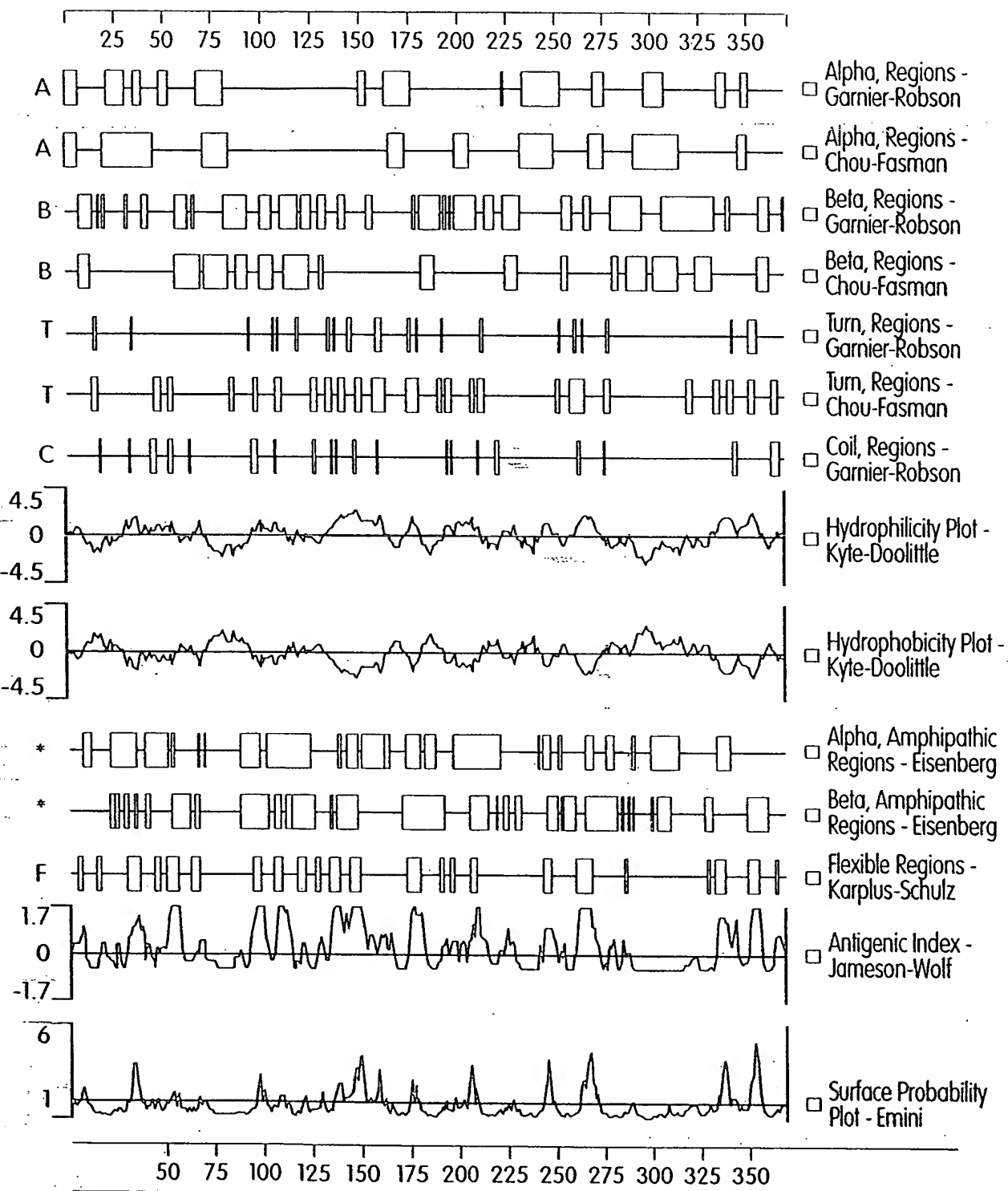
Fig. 11A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CNIM
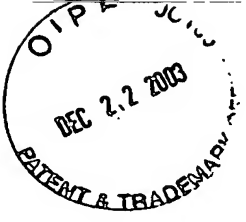Sheet 16 of 43
16/43

```
Y   V   G   N   V   A   W   M   H   V   L   A   A   R   E   L   E   Q   R   A    249
TAT GTG GGC AAT GTT GCC TGG ATG CAC GTG CTG GCA GCC CGG GAG CTG GAG CAG CGG GCA  747

A   L   M   G   G   Q   V   Y   F   C   Y   D   G   S   P   Y   R   S   Y   E    269
GCC CTG ATG GGC GGC CAG GTA TAC TTC TGC TAC GAT GGA TCA CCC TAC AGG AGC TAC GAG  807

D   F   N   M   E   F   L   G   P   C   G   L   R   L   V   G   A   R   P   L    289
GAT TTC AAC ATG GAG TTC CTG GGC CCC TGC GGA CTG CGG CTG GTG GGC GCC CGC CCA TTG  867

L   P   Y   W   L   L   V   F   L   A   A   L   N   A   L   L   Q   W   L   L    309
CTG CCC TAC TGG CTG CTG GTG TTC CTG GCT GCC CTC AAT GCC CTG CTG CAG TGG CTG CTG  927

R   P   L   V   L   Y   A   P   L   L   N   P   Y   T   L   A   V   A   N   T    329
CGG CCA CTG GTG CTC TAC GCA CCC CTG CTG AAC CCC TAC ACG CTG GCC GTG GCC AAC ACC  987

T   F   T   V   S   T   D   K   A   Q   R   H   F   G   Y   E   P   L   F   S    349
ACC TTC ACC GTC AGC ACC GAC AAG GCT CAG CGC CAT TTC GGC TAT GAG CCC CTG TTC TCG 1047

W   E   D   S   R   T   R   T   I   L   W   V   Q   A   A   T   G   S   A   Q    369
TGG GAG GAT AGC CGG ACC CGC ACC ATT CTC TGG GTA CAG GCC GCT ACG GGT TCA GCC CAG 1107

*                                                                                370
TGA                                                                             1110
```

CGGTGGGGCTGGGGCCTGGAGGCCCAGATACAGCACATCCACCCAGGTCCCGAGCCCTCACACCCTGGACGGGAAGGGA

CAGCTGCATTCCAGAGCAGGAGGCAGGGCTCTGGGGCCAGAATGGCTGTCCTTGTCGTAGAGCCCTCCACATTTTCTTT

TTCTTTTTTGAGACAGGGTCTTGCTCTGTCACCCAGACTGGAATGCAAGTGGTGTGANTCATAAGCTCACTNGMACCCT

YAANCCTTCTGGGTTCAAGCAATCCTTNCTNGCCTYAANCCTTCTNGAACAAGCTTGGGANCCACAGGTGCACGCCANC

CACANCCTGGCTTTTTTTT

Fig. 11B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
 THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 17 of 43

17/43

Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte-Doolittle
Hydrophobicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenberg
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini

Fig. 12

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
10	26	out-->ins	1.0
73	90	ins-->out	2.0
289	305	out-->ins	3.0
312	333	ins-->out	1.2

>25964
MADSAQAQKLVYLVTGGCGFLGEHVVRMLLQREPRLGELRVFDQHLGPWLEELKTGPVRV
TAIQGDVTQAHEVAAAVAGAHVVIHTAGLVDVFGRASPKTIHEVNVQGTRNVIEACVQTG
TRFLVYTSSMEVVGPNTKGHPFYRGNEDTPYEAVHRHPYPCSKALAEWLVLEANGRKVRG
GLPLVTCALRPTGIYGEGHQIMRDFYRQGLRLGGWLFRAIPASVEHGRVYVGNVAWMHVL
AARELEQRAALMGGQVYFCYDGSPYRSYEDFNMEFLGPCGLRLVGARPLLPYWLLVFLAA
LNALLQWLLRPLVLYAPLLNPYTLAVANTTFTVSTDKAQRHFGYEPLFSWEDSRTRTILW
VQAATGSAQ

Fig. 13

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 19 of 43
19/43

Protein Family / Domain Matches, HMMer Version 2

```
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:            /prod/ddm/seqanal/PFAM/pfam5.0/Pfam
Sequence file:       /prod/ddm/wspace/orfanal/oa-script.9289.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  25964


Scores for sequence family classification (score includes all domains):
Model           Description                              Score    E-value  N
--------        -----------                              -----    -------  ---
3Beta HSD       3-beta hydroxysteroid dehydrogenase/iso  676.9    1e-199   1
S-AdoMet synt   S-adenosylmethionine synthetase            1.8      0.78   1
adh_short       short chain dehydrogenase                -48.6      0.022  1
Epimerase       NAD dependent epimerase/dehydratase fam  -148.0     0.0016 1


Parsed for domains:
Model           Domain  seq-f  seq-t    hmm-f hmm-t       score   E-value
--------        ------- ----------      ----- -----       -----   -------
adh_short        1/1      10    197 ..     1   203  []    -48.6    0.022
S-AdoMet_synt    1/1     341    351 ..   365   376  .]      1.8    0.78
3Beta_HSD        1/1       1    365 [.     1   425  []    676.9    1e-199
Epimerase        1/1      12    365 ..     1   359  []   -148.0    0.0016


Alignments of top-scoring domains:
adh_short: domain 1 of 1, from 10 to 197: score -48.6, E = 0.022
                 *->KvaLvTGassGIGlaiAkrLakeGakVvvadrneeklekGavalelk
                    v LvTG+++ +G +++ L+ +      ++ ++  +   G +++elk
        25964   10   LVYLVTGGCGFLGEHVVRMLLQR--EPRLGELRVFDQHLGPWLEELK 54


                 elGgndkdralaiqlDvtdeesv.aaveqaverlGrlDvLVNNAGgiill
                    +     r+ aiq+Dvt++ +v aav+ a        +v++ AG +
        25964   55 TGPV----RVTAIQGDVTQAHEVaAAVAGA-------HVVIHTAG--L-- 89


                 rpgpfaelsrtmeedwdrvidvNltgvflltravlplmamkkrggGrIvN
                    +  f + s  ++    +++vN+ g    tr v++  a ++ g    v
        25964   90 -VDVFGRAS---PK---TIHEVNVQG----TRNVIE--ACVQTGTRFLVY 126


                 iSSvaGrke..............g.glvgvpggsaYsASKaAvigltrs
                 +SS     +e ++++++++  +++ + +    ++ +Y +SKa   l++
        25964  127 TSS----MEvvgpntkghpfyrgnEdTPYEAVHRHPYPCSKA----LAEW 168


                 LAlElaphgIr......VnavaPGgvdTd<-*
                 L lE +++r++ +    a  P g++ +
        25964  169 LVLEANGRKVRgglplvTCALRPTGIYGE     197


S-AdoMet_synt: domain 1 of 1, from 341 to 351: score 1.8, E = 0.78
                  *->HFGreevdFpWE<-*
                     HFG e   F+WE
        25964  341   HFGYEP-LFSWE     351
```

Fig. 14A

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 20 of 43
20/43

```
3Beta_HSD: domain 1 of 1, from 1 to 365: score 676.9, E = 1e-199
                 *->elseseldmaglsclVTGGgGFlGrhIVreLlregeslqevRvfDlrf
                    +++s++    l++lVTGG+GFlG+h Vr+Ll+++++l e+RvfD +
     25964     1      -MADSAQAQKLVYLVTGGCGFLGEHVVRMLLQREPRLGELRVFDQHL 46


                 spelde.dssklqvitkikyieGDvtDkqdlaaAlqgiSCCTLLDmTLmD
                    +p+l+e +++++ v+      +i+GDvt+++++aaA++g+
     25964    47 GPWLEElKTGPVRVT----AIQGDVTQAHEVAAAVAGA------------ 80


                 dvvIHtAaiiDvfGelrvsGSDLSFGVTVLFLAVTEGSYVVFYmGATDLR
                    +vvIHtA+++DvfG
     25964    81 HVVIHTAGLVDVFG------------------------------------ 94


                 kasrdrimkVNVkGTqnvldACveaGVrvlVYTSSmeVVGpNsrGqpivN
                    as+. +i++VNV+GT+nv++ACv++G+r+lVYTSSmeVVGpN +G+p+++
     25964    95 RASPKTIHEVNVQGTRNVIEACVQTGTRFLVYTSSMEVVGPNTKGHPFYR 144


                 GdEttpYestDDhqdaYpeSKalAEklVLkANGsmlknGgrLyTCALRPa
                    G+E+tpYe++   h+++Yp+SKalAE lVL+ANG+ +++G  L+TCALRP+
     25964   145 GNEDTPYEAV--HRHPYPCSKALAEWLVLEANGRKVRGGLPLVTCALRPT 192


                 gIfGeGdqflvpflrqlvknGlakfriGdknalsdrVYVgNVAwAHILAA
                    gI+GeG q + +f+rq +++G+ +fr      ++ + rVYVgNCAw+H+LAA
     25964   193 GIYGEGHQIMRDFYRQGLRLGGWLFRAIPASVEHGRVYVGNVAWMHVLAA 242


                 raLqdpkkGREGassiaGqaYFIsDdsPvnSYddFnrtllkalGlrlpst
                    r+L+++       a+ + Gq+YF++D+sP++SY+dFn+++l ++Glrl +
     25964   243 RELEQR------AALMGGQVYFCYDGSPYRSYEDFNMEFLGPCGLRLVGA 286


                 w.rlPlpllyvlaylnellswLLrklalrYtPllnpytvtlanttFtfst
                    ++lP++ll++la+ln+ll+wLLr+l + Y Pllnpyt+++anttFt+st
     25964   287 RpLLPYWLLVFLAALNALLQWLLRPL-VLYAPLLNPYTLAVANTTFTVST 335


                 nKAkkdLGYePlvtwEEarakTieWiqele<-*
                    +KA+++GYePl++wE +r +Ti+W+q+
     25964   336 DKAQRHFGYEPLFSWEDSRTRTILWVQAAT        365
```

Fig. 14B-1
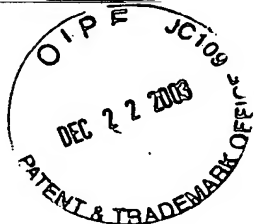
Epimerase: domain 1 of 1, from 12 to 365: score -148.0, E = 0.0016

```
              *->ILVTGGAGFIGShlvreLlnn...ygddkVvvLDnLtdyYqyagnea
                 +LVTGG GF G h+vr Li+ +++   g  +V    +      + +
    25964    12   YLVTGGCGFLGEHVVRMLLQReprLGELRV-----FD----QHLGPW 49


                   rlevvegnprytFvkGDIcDrdlldkvfaehqpDaViHfAAeshV.drSi
                   ++e + g  r+t ++GD+ + +   ++a    +ViH A++  V +r
    25964    50   LEELKTGPVRVTAIQGDVTQAHEVAAAVAGA--HVVIHTAGLVDVfGR-- 95


                   ekPlayidtNvvGTltLLEaaRnYWsaLdetkagvkkfvfsSTdeVYGdl
                    P + + Nv GT + +Ea+           g  +v+ S+ eV G +
    25964    96   ASPKTIHEVNVQGTRNVIEACV---------QTGTRFLVYTSSMEVVGPN 136


                   esiPisaF...tEdtPynPs..SPYgaSKassEllvrayhraygLpaiiL
                   ++ + F ++ EdtPy   ++ PY  SKa  E  lv   +
    25964    137  TKGHP--FyrgNEDTPYEAVhrHPYPCSKALAEWLVLEAN---------- 174


                   RyFNvYGpyqsgriGedpngfpekLIPliiqnalgkgeplpvYGdDYpTp
                        G+    g+            +Pl+ + al    p  +YG
    25964    175  ------GRKVRGG-----------LPLV-TCALR---PTGIYG------ 196


                   DGtqv.RDw....................ihVeDharANhllaltkg
                   +G q+ RD+ +++ + ++   +  + + +++++++V ++a   h+la +++
    25964    197  EGHQImRDFyrqglrlggwlfraipasvehgrVYVGNVAWM-HVLAAREL 245


                   ...............................raGkgsevYNiGg
                   +++     +++     ++++ ++ ++ + +  ++ + + +G
    25964    246  eqraalmggqvyfcydgspyrsyedfnmeflgpcglrLVG---------- 285


                   gneysnlEvVeaIekllgelaPekphvkakedpatfvddRpGddarya..
                    +  + + + +++++l+ 1           + ++ +++    ++++a
    25964    286  -ARPLLPYWLLVFLAALNALLQWL-----LRPLVLiYAPLLN--PYTLAva 327


                   .....aDasKikreLGWkPevtnleeGladTvnWylene<-*
                   +++ +++ +K++I  G++P + e+  +T+ W +
    25964    328  nttftVSTDKAQRHFGYEPLFS-WEDSRTRTILWVQAAT      365
```

Fig 14B-2

ProdomId	Start	End	Description	Score
View Prodom 1280 [Boxer ▼] [Showing match ▼] [Go!]	11	362	p99.2 (39) 3BHS(5) 3BH1(4) 3BH2(3) // DEHYDROGENASE STEROID BETA-HYDROXYSTEROID 3BETA-HSD DEHYDROGENASE/DELTA 5-->4-ISOMERASE INCLUDES: PROGESTERONE 3-BETA-HYDROXY-DELTA5-STEROID 3-BETA-HYDROXY-5-ENE	395
ProdomId	Start	End	Description	Score

View Prodom 1280 [Boxer ▼] [Showing match ▼] [Go!]

>1280 p99.2 (39) 3BHS(5) 3BH1(4) 3BH2(3) // DEHYDROGENASE STEROID
    BETA-HYDROXYSTEROID 3BETA-HSD DEHYDROGENASE/DELTA 5-->4-ISOMERASE
    INCLUDES: PROGESTERONE 3-BETA-HYDROXY-DELTA5-STEROID 3-BETA-HYDROXY-5-ENE
    Length = 416

 Score = 395 (144.1 bits), Expect = 3.2e-42, Sum P(2) = 3.2e-42
 Identities = 99/268 (36%), Positives = 134/268 (50%)

Query: 102 HEVNVQGTRNVIEACVQTGTRFLVYTSSMEVVGPNTKGHPFYRGNEDTPYEAVHRHPYPC 161
            ++ NVQGTRN+IE C   RF   MEV GPN+   G+E+ +E+   +PYP
Sbjct: 157 YKFNVQGTRNLIEKC-----RFF---GVMEVAGPNSYKEIILNGHEEEHHESTWPNPYPY 208

Query: 162 -SKALAEWLVLEANGRKVRGGLPLVTCALRPTGIYGEGHQIMRDFYRQGLRLGGWLFRAI 220
             SK +AE VL ANG ++ G L TCALRP IYGEG + +   Q L+ GG +FR
Sbjct: 209 YSKKMAEKAVLAANGSMLKNGGTLYTCALRPMYIYGEGDKFLSPMIVQALKNGGIMFRVG 268

Query: 221 PASVEHGRVYVGNVAWMHVXXXXXXXXXXXXX--MGGQVYFCYDGSPYRSYEDFNMEFLGP 278
                 VYVGNVAW H+   + GQ Y+ D +P++SY+D N
Sbjct: 269 GKFSVANPVYVGNVAWAHILAARGLQDPKKSPNIQGQFYYISDDTPHQSYDDLNYTLSKE 328

Query: 279 CGLRLVGARPLLP----YWXXXXXXXXXXXXXXXXXXXXXXXXXXXXNPYTLAVANTTFTVS 334
            GLRL ++ LP   YW   N + + ++NTTFT S
Sbjct: 329 WGLRLDSSKWRLPLPLLYWLAFLLEMVSFLLRPISYNYQPPF---NRHLVTLSNTTFTFS 385

Query: 335 TDKAQRHFGYEPLFSWEDSRTRTILWVQ 362
            KAQR GYEPL SWE+++ +T W++
Sbjct: 386 YKKAQRDLGYEPLVSWEEAKQKTSEWIE 413

 Score = 65 (27.9 bits), Expect = 3.2e-42, Sum P(2) = 3.2e-42
 Identities = 11/23 (47%), Positives = 17/23 (73%)

Query: 11 VYLVTGGCGFLGEHVVRMLLQRE 33
            VY VTGG FLG ++V++L+ +
Sbjct: 14 VYAVTGGAEFLGRYIVKLLISAD 36

Fig. 15

Input file Fbh21686F1.seq; Output File 21686.trans
Sequence length 1209

```
                                                     M   S   L   R       4
CCCACGCGTCCGCCCACGCGTCCGCGGACGCGTGGGCGGACGCGTGGGCGCCCGCCTCGA ATG TCC CTG AGA    12

  P   R   R   A   C   A   Q   L   L   W   H   P   A   A   G   M   A   S   W   A   24
CCC AGA AGG GCC TGC GCT CAG CTG CTC TGG CAC CCC GCT GCA GGG ATG GCC TCC TGG GCT   72

  K   G   R   S   Y   L   A   P   G   L   L   Q   G   Q   V   A   I   V   T   G   44
AAG GGC AGG AGC TAC CTG GCG CCT GGT TTG CTG CAG GGC CAA GTG GCC ATC GTC ACC GGC   132

  G   A   T   G   I   G   K   A   I   V   K   E   L   L   E   L   G   S   N   V   64
GGG GCC ACG GGC ATC GGA AAA GCC ATC GTG AAG GAG CTC CTG GAG CTG GGG AGT AAT GTG   192

  V   I   A   S   R   K   L   E   R   L   K   S   A   A   D   E   L   Q   A   N   84
GTC ATT GCA TCC CGT AAG TTG GAG AGA TTG AAG TCT GCG GCA GAT GAA CTG CAG GCC AAC   252

  L   P   P   T   K   Q   A   R   V   I   P   I   Q   C   N   I   R   N   E   E   104
CTA CCT CCC ACA AAG CAG GCA CGA GTC ATT CCC ATA CAA TGC AAC ATC CGG AAT GAG GAG   312

  E   V   N   N   L   V   K   S   T   L   D   T   F   G   K   I   N   F   L   V   124
GAG GTG AAT AAT TTG GTC AAA TCT ACC TTA GAT ACT TTT GGT AAG ATC AAT TTC TTG GTG   372

  N   N   G   G   G   Q   F   L   S   P   A   E   H   I   S   S   K   G   W   H   144
AAC AAT GGA GGA GGC CAG TTT CTT TCC CCT GCT GAA CAC ATC AGT TCT AAG GGA TGG CAC   432

  A   V   L   E   T   N   L   T   G   T   F   Y   M   C   K   A   V   Y   S   S   164
GCT GTG CTT GAG ACC AAC CTG ACG GGT ACC TTC TAC ATG TGC AAA GCA GTT TAC AGC TCC   492

  W   M   K   E   H   G   G   S   I   V   N   I   I   V   P   T   K   A   G   F   184
TGG ATG AAA GAG CAT GGA GGA TCT ATC GTC AAT ATC ATT GTC CCT ACT AAA GCT GGA TTT   552

  P   L   A   V   H   S   G   A   A   R   A   G   V   Y   N   L   T   K   S   L   204
CCA TTA GCT GTG CAT TCT GGA GCT GCA AGA GCA GGT GTT TAC AAC CTC ACC AAA TCT TTA   612

  A   L   E   W   A   C   S   G   I   R   I   N   C   V   A   P   G   V   I   Y   224
GCT TTG GAA TGG GCC TGC AGT GGA ATA CGG ATC AAT TGT GTT GCC CCT GGA GTT ATT TAT   672

  S   Q   T   A   V   E   N   Y   G   S   W   G   Q   S   F   F   E   G   S   F   244
TCC CAG ACT GCT GTG GAG AAC TAT GGT TCC TGG GGA CAA AGC TTC TTT GAA GGG TCT TTT   732

  Q   K   I   P   A   K   R   I   G   V   P   E   E   V   S   S   V   V   C   F   264
CAG AAA ATC CCC GCT AAA CGA ATT GGT GTT CCT GAG GAG GTC TCC TCT GTG GTC TGC TTC   792

  L   L   S   P   A   A   S   F   I   T   G   Q   S   V   D   V   D   G   G   R   284
CTA CTG TCT CCT GCA GCT TCC TTC ATC ACT GGA CAG TCG GTG GAT GTG GAT GGG GGC CGG   852

  S   L   Y   T   H   S   Y   E   V   P   D   H   D   N   W   P   K   G   A   G   304
AGT CTC TAT ACT CAC TCG TAT GAG GTA CCA GAT CAT GAC AAC TGG CCC AAG GGA GCA GGG   912

  D   L   S   V   V   K   K   M   K   E   T   L   K   E   K   A   K   L   *       323
GAC CTT TCT GTT GTC AAA AAG ATG AAG GAG ACC TTA AAG GAG AAA GCT AAG CTC TGA       969
```

GCTGAGGAAACAAGGTGTCCTCCATCCCCAGTGCCTTCACATCTTGAGGATATGCTTCTGTACTTTTTAAAAGCTTATA

GTTGGTATGGAAAACATTTTTCTTATTTTTAAGTGTTATTAATTATATCTATGGAAAAACTATTCCTGAAATATATACA

GTCTTATGTCCCAAAAAAAAAA

Fig. 16

CLUSTAL W (1.74) multiple sequence alignment

5052204_SDR_rat --------------------MGSWKSGQSYLAAGLLQNQVAVVTGGATGIGKAISRELLHL
21686   MSLRPRRACAQLLWHPAAGMASWAKGRSYLAPGLLQGQVAIVTGGATGIGKAIVKELLEL
                 *.** .*:****.****.***:************* .:***.*

5052204_SDR_rat GCNVVIASRKLDRLTAAVDELRASQPPSSSTQVTAIQCNIRKEEEVNNLVKSTLAKYGKI
21686   GSNVVIASRKLERLKSAADELQANLPPTKQARVIPIQCNIRNEEEVNNLVKSTLDTFGKI
                 *.*********:**.:*.***:*. **:..::* .******:*********** .:***

5052204_SDR_rat NFLVNNAGGQFMAPAEDITAKGWQAVIETNLTGTFYMCKAVYNSWMKDHGGSIVNIIVLL
21686   NFLVNNGGGQFLSPAEHISSKGWHAVLETNLTGTFYMCKAVYSSWMKEHGGSIVNIIVPT
                 ******.****::***.*::***:**:****************.****:**********

5052204_SDR_rat NNGFPTAAHSGAARAGVYNLTKTMALTWASSGVRINCVAPGTIYSQTAVDNYGELGQTMF
21686   KAGFPLAVHSGAARAGVYNLTKSLALEWACSGIRINCVAPGVIYSQTAVENYGSWGQSFF
                 : *** *.**************:.:** **.**:******* *******.***. **::*

5052204_SDR_rat EMAFENIPAKRVGLPEEISPLVCFLLSPAASFITCQLINVDGGQALYTRNFTIPDHDNWP
21686   EGSFQKIPAKRIGVPEEVSSVVCFLLSPAASFITGQSVDVDGGRSLYTHSYEVPDHDNWP
                 * :*::*****:*:***:*.:.:************ ::****::***:.: :*******

5052204_SDR_rat VGAGDSSFIKKVKESLKKQARL
21686   KGAGDLSVVKKMKETLKEKAKL
                 **** *.:**:**.**::*.*
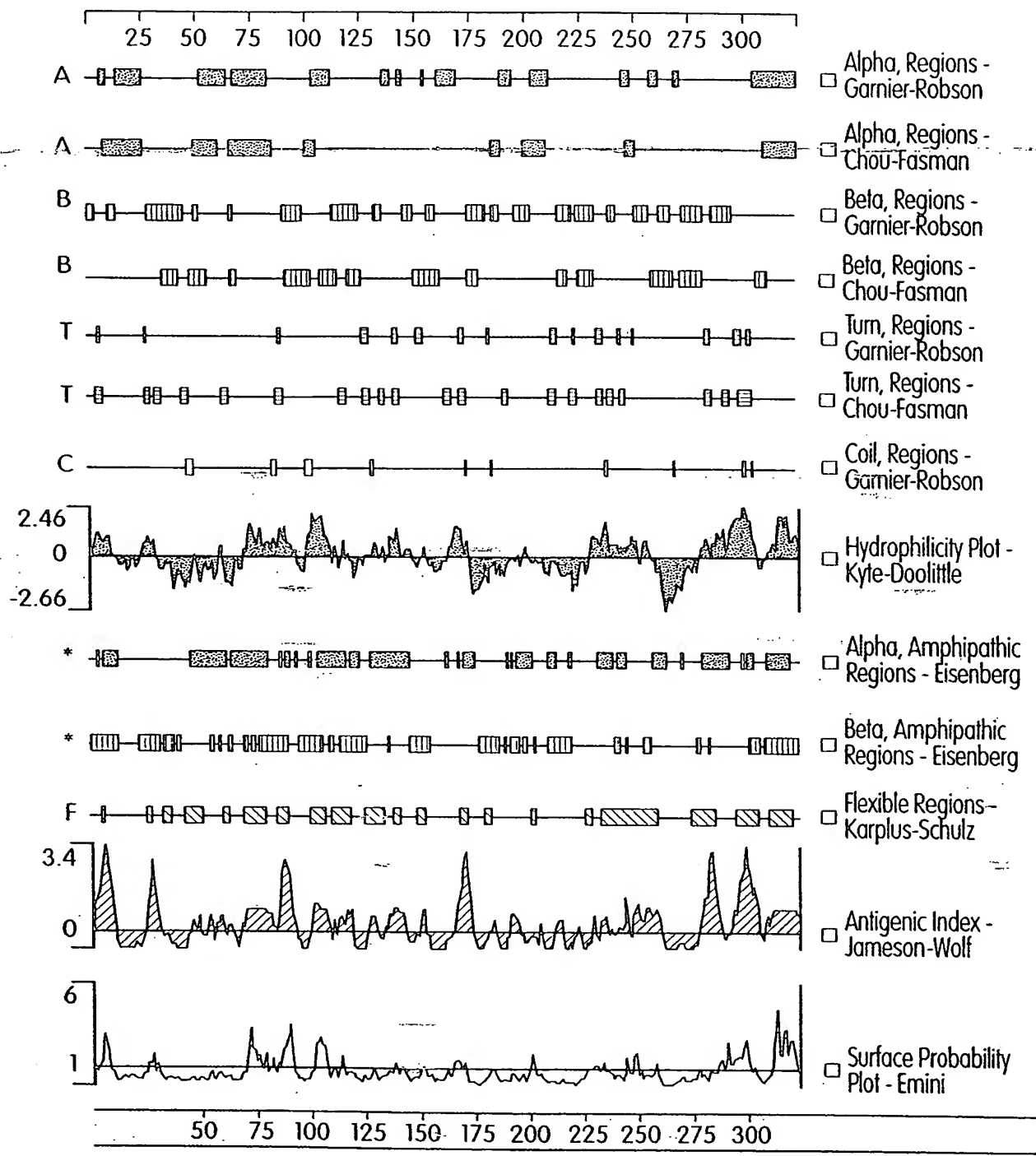
Fig. 18

Signal Peptide Predictions for 21686

Method	Predict	Score	Mat@
Signal (eukaryote)	MAYBE		20

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
29	50	ins-->out	0.9
170	188	out-->ins	0.2
208	224	ins-->out	0.6
258	275	out-->ins	2.6

```
>21686
MSLRPRRACAQLLWHPAAGMASWAKGRAYLAPGLLQGQVAIVTGGATGIGKAIVKELLEL
GSNVVIASRKLERLKSAADELQANLPPTKQARVIPIQCNIRNEEEVNNLVKSTLDTFGKI
NFLVNNGGGQFLSPAEHISSKGWHAVLETNLTGTFYMCKAVYSSWMKEHGGSIVNIIVPT
KAGFPLAVHSGAARAGVYNLTKSLALEWACSGIRINCVAPGVIYSQTAVENYGSWGQSFF
EGSFQKIPAKRIGVPEEVSSVVCFLLSPAASFITGQSVDVDGGRSLYTHSYEVPDHDNWP
KGAGDLSVVKKMKETLKEKAKL
```

Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
10	31	ins-->out	0.9
151	169	out-->ins	0.2
189	205	ins-->out	0.6
239	256	out-->ins	2.6

```
>21686_mature
MASWAKGRSYLAPGLLQGQVAIVTGGATGIGKAIVKELLELGSNVVIASRKLERLKSAAD
ELQANLPPTKQARVIPIQCNIRNEEEVNNLVKSTLDTFGKINFLVNNGGGQFLSPAEHIS
SKGWHAVLETNLTGTFYMCKAVYSSWMKEHGGSIVNIIVPTKAGFPLAVHSGAARAGNYN
LTKSLALEWACSGIRINCVAPGVIYSQTAVENYGSWGQSFFEGSFQKIPAKRIGVPEEVS
SVVCFLLSPAASFITGQSVDVDGGRSLYTHSYEVPDHDNWPKGAGDLSVVKKMKETLKEK
AKL
```

Fig. 19

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 27 of 43

27/43

Protein Family / Domain Matches, HMMer Version 2

```
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:              /prod/ddm/seqanal/PFAM/pfam4.4/Pfam
Sequence file:         /prod/ddm/wspace/orfanal/oa-script.19160.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:   21686


Scores for sequence family classification (score includes all domains):
Model          Description                               Score   E-value  N
--------       -----------                               -----   -------  ---
adh_short      short chain dehydrogenase                 162.5   7.3e-45  1
adh_short_C2   short chain dehydrogenase/reductase C-te   47.2   3.7e-10  1


Parsed for domains:
Model          Domain   seq-f  seq-t   hmm-f  hmm-t     score  E-value
--------       -------  -----  -----   -----  -----     -----  -------
adh_short      1/1         38    226 ..     1    203 [] 162.5  7.3e-45
adh_short_C2   1/1        250    280 ..     1     31 []  47.2  3.7e-10


Alignments of top-scoring domains:
adh_short: domain 1 of 1, from 38 to 226: score 162.5, E = 7.3e-45
                 *->KvaLvTGassGIGlaiAkrLakeGakVvvadrneeklekGavakelk
                    +va+vTG++ GIG+ai+k+L++ G +Vv+a r  e+l    ++++
     21686     38    QVAIVTGGATGIGKAIVKELLELGSNVVIASRKLERL-----KSAAD 79


                 elGgnd....kdralaiqlDvtdeesv.aaveqavarlGrlDvLVNNAGg
                 el +n+++++  r+++iq++++ ee+v+++v+  ++ +G+++ LVNN Gg
     21686     80 ELQANLpptkQARVIPIQCNIRNEEEVnNLVKSTLBiFGKINFLVNNGGG 129


                 .iillrpgpfaelsrtmeedwdrvidvNltgvflltravlplmamkkrgg
                 +++      p++ +s     + w +v+++Nltg+f++++av     +k +g
     21686    130 qFL----SPAEHIS---SKGWHAVLETNLTGTFYMCKAVYS--SWMKEHG 170


                 GrIvNiSSvaGrkegglvgvpggsaYsASKaAvigltrsLAlElaphgIr
                 G+IvNi           + g+p +++ +A+ a+v lt+sLAlE+a  gIr
     21686    171 GSIVNIIV-PT-----KAGFPLAVHSGAARAGVYNLTKSLALEWACSGIR 214


                 VnavaPGgvdTd<-*
                 +n+vaPG ++ +
     21686    215 INCVAPGVIYSQ     226


adh_short_C2: domain 1 of 1, from 250 to 280: score 47.2, E = 3.7e-10
                 *->gRlGePeEiAnavvFLASdaAsYiTGqtlvV<-*
                    +R G PeE++++v FL S+aAs+iTGq + V
     21686    250  KRIGVPEEVSSVVCFLLSPAASFITGQSVDV     280
```
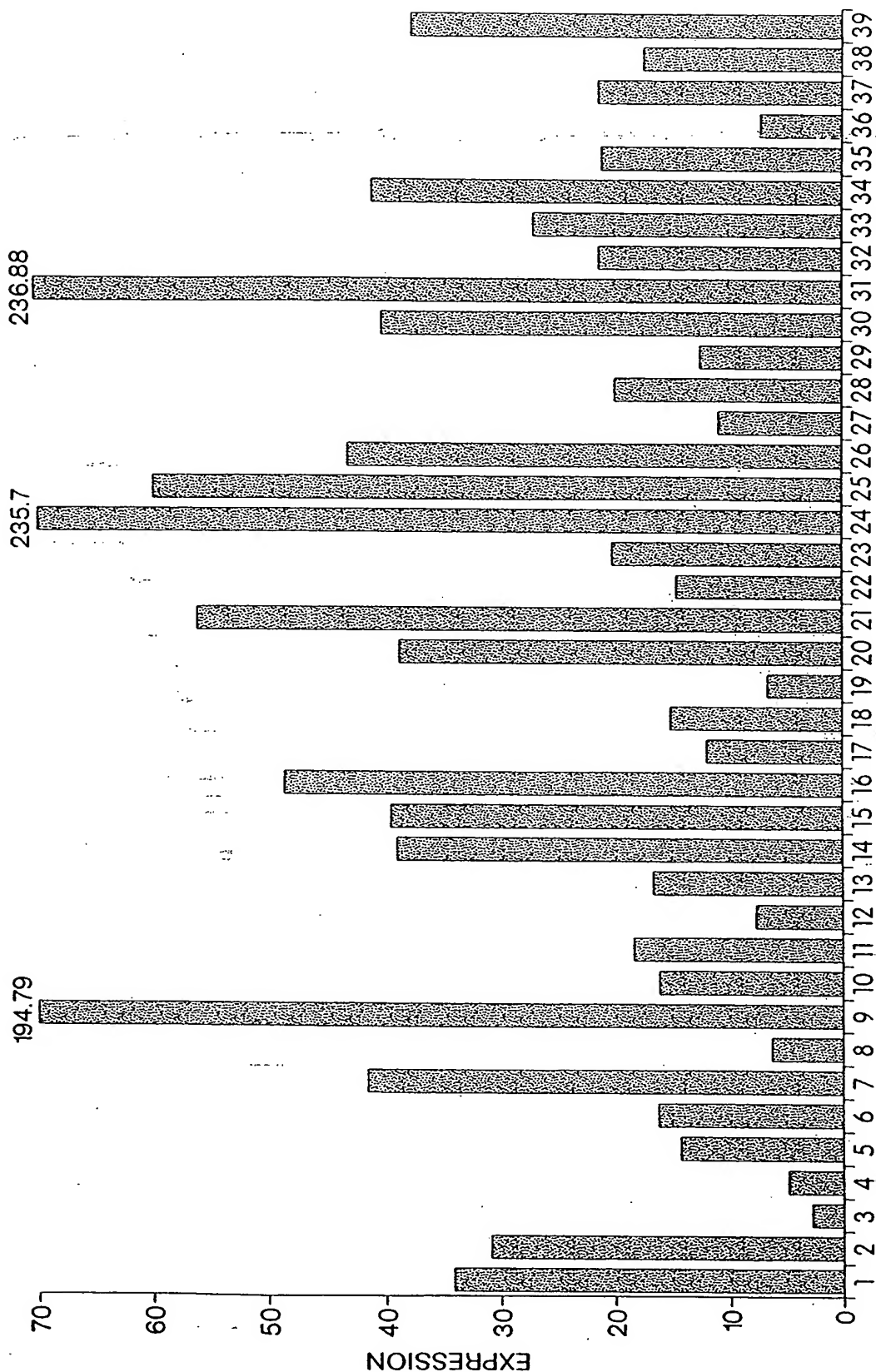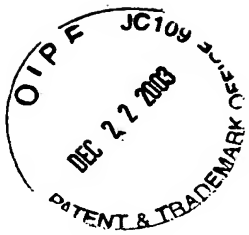
Fig. 20

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 28 of 43
28/43

ProDom Matches

ProdomId		Start	End	Description	Score
View Prodom 121622 [Boxer ▼] [Showing match ▼] [Go!]		29	82	p99.2 (1) YSO5_CAEEL // HYPOTHETICAL 98.0 KD PROTEIN F56D1.5 IN CHROMOSOME II TRANSMEMBRANE	70
View Prodom 95301 [Boxer ▼] [Showing match ▼] [Go!]		35	82	p99.2 (1) O27957_ARCFU // SHIKIMATE 5-DEHYDROGENASE AROE HYPOTHETICAL PROTEIN	86
View Prodom 11 [Boxer ▼] [Showing match ▼] [Go!]		37	231	p99.2 (1078) ADH(34) GALE(20) FABG(13) // OXIDOREDUCTASE PROTEIN DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIS SYNTHASE ALCOHOL PUTATIVE	157
View Prodom 73753 [Boxer ▼] [Showing match ▼] [Go!]		237	286	p99.2 (1) P71079_BACSU // UNIDENTFIED DEHYDROGENASE	84
View Prodom 77223 [Boxer ▼] [Showing match ▼] [Go!]		243	287	p99.2 (1) O07882_STAXY // GLUCOSE-1-DEHYDROGENASE	92
ProdomId		Start	End	Description	Score

View Prodom 11 [Boxer ▼] [Showing match ▼] [Go!]

```
>11 p99.2 (1078) ADH(34) GALE(20) FABG(13)  // OXIDOREDUCTASE PROTEIN
    DEHYDROGENASE NAD REDUCTASE NADP BIOSYNTHESIA SYNTHASE ALCOHOL PUTATIVE
    Length = 269
```

Score = 157 (60.3 bits), Expect = 1.2e-09, P = 1.2e-09
Identities = 64/213 (30%), Positives = 106/213 (49%)

```
Query:    51 KAIVKELLELGSNVVIASRKLERLKSAADELQANLPPTKQA---RVIPIQCNIRNEEEVN 107
             K +V        S    AS+  E     +  +A+ T QA    V  + C++ + E+V
Sbjct:    35 KVVVVSATSEESESTEASK--ESAMEVSKAVNAEVSATMQAVGVTVTKVTCDVADVEDVE 92

Query:   108 NLVKSTLDTF----GKINFLVNNGGGQFLSP---AEHISSKG-----WHAVLETNLTGTF 155
             LV++ ++ F    GKI+ LVNN G  ++P   AE ++ +       W  V+E N+TGTF
Sbjct:    93 KLVETVVEEFSGIHGKIDVLVNNAG--VMAPKAVAESMTEETSDDEEWEEVIEVNVTGTF 150

Query:   156 YMCKAVYSSWMK--------EHGGSIVNI--IVPTKAGFP--LAVHSGAARAGVYNLTKS 203
             + +A + K          G+IVN+ +  + G P   A +S A++A V + TKS
Sbjct:   151 NLTQAALPAMKKFSDAAAKKRFVGTIVNVASVAGSTMGSPGSQAAYS-ASKAAVESFTKS 209

Query:   204 LALE---WACSG--IRINCVAPGVIYSQTAVEN 231
             LA+E   ++ S  +R+N VAPG + +  A+E+
Sbjct:   210 LAMELSPYSASVAMVRVNAVAPGYVETD-ALES 241
```

Score = 103 (41.3 bits), Expect = 0.0021, Sum P(2) = 0.0021
Identities = 32/100 (32%), Positives = 54/100 (54%)

```
Query:    37 GQVAIVTGGA--TGIGKAIVKELLELGSNVVIASRKLERLKS--AADE--------LQAN 84
             G+  +VTGG+  +GIG AI ++L E G+ VV+ S    E  +S A+ E        + A
Sbjct:     7 GKTVLVTGGSGFSGIGLAIARQLAEEGAKVVVVSATSEESESTEASKESAMEVSKAVNAE 66

Query:    85 LPPTKQA---RVIPIQCNIRNEEEVNNLVKSTLDTFGKIN 121
             +  T QA    V  + C++ + E+V LV++ ++ F  I+
Sbjct:    67 VSATMQAVGVTVTKVTCDVADVEDVEKLVETVVEEFSGIH 106
```

Score = 37 (18.1 bits), Expect = 0.0021, Sum P(2) = 0.0021
Identities = 9/23 (39%), Positives = 13/23 (56%)

```
Query:   205 ALEWACSGIRINCVAPGVIYSQT 227
             ALE A +G+ +  V PG +   T
Sbjct:   238 ALESATNGLSVVTVRPGNVRVNT 260
```

Fig. 21A

View Prodom 77223 [Boxer ▼] [Showing match ▼] [Go!]

>77223 p99.2 (1) 007882_STAXY // GLUCOSE-1-DEHYDROGENASE
        Length = 67

Score = 92 (37.4 bits), Expect = 0.00031, P = 0.00031
 Identities = 19/45 (42%), Positives = 29/45 (64%)

Query: 243 SFQKIPAKRIGVPEEVSSVVCFLLSPAASFITGQSVDVDGGRSLY 287
              + + IPAK IG ++V++V FL S A +I G ++ VDGG + Y
Sbjct: 15 TLEMIPAKEIGFADQVANVARFLCSDLADYIHGTTIYVDGGMTNY 59

View Prodom 95301 [Boxer ▼] [Showing match ▼] [Go!]

>95301 p99.2 (1) O27957_ARCFU // SHIKIMATE 5-DEHYDROGENASE
AROE HYPOTHETICAL
        PROTEIN
        Length = 108

 Score = 86 (35.3 bits), Expect = 0.0014, P = 0.0014
 Identities = 20/48 (41%), Positives = 31/48 (64%)

Query: 35 LQGQVAIVTGGATGIGKAIVKELLELGSNVVIASRKLERLKSAADELQ 82
              L G+ A+V G A G GKA   LL++GS V++A+R E+ + A + L+
Sbjct: 10 LGGKTALVVG-AGGAGKAAALALLDMGSTVIVANRTEEKGREAVEMLR 56

View Prodom 73753 [Boxer ▼] [Showing match ▼] [Go!]

>73753 p99.2 (1) P71079_BACSU // UNIDENTFIED DEHYDROGENASE
        Length = 60

 Score = 84 (34.6 bits), Expect = 0.0023, P = 0.0023
 Identities = 20/50 (40%), Positives = 29/50 (58%)

Query: 237 QSFFEGSFQKIPAKRIGVPEEVSSVVCFLLSPAASFITGQSVDVDGGRSL 286
              + E + Q PA R+   +++   V FL+S A I GQ++ VDGGRSL
Sbjct: 9 EDLLEDARQNTPAGRMVEIKDMVDTVEFLVSSKADMIRGQTIIVDGGRSL 58

View Prodom 121622 [Boxer ▼] [Showing match ▼] [Go!]

>121622 p99.2 (1) YSO5_CAEEL // HYPOTHETICAL 98.0 KD PROTEIN F56D1.5 IN
        CHROMOSOME II TRANSMEMBRANE
        Length = 194

 Score = 70 (29.7 bits), Expect = 7.6, P = 1.0
 Identities = 20/57 (35%), Positives = 29/57 (50%)

Query: 29 YLAPG;;QGQV--AIVTGGATGIGKAIVKELLELG-SNVVIASRKLERLKSAADELQ 82
              + P L Q Q   +V+GG GIGKA   EL + G V+ R ++L S E++
Sbjct: 62 FYKPNLEQYQHRWTVVSGGTDGIGKAYTLELAKRGLRKFVLIGRNPKKLDSVKSEIE 118
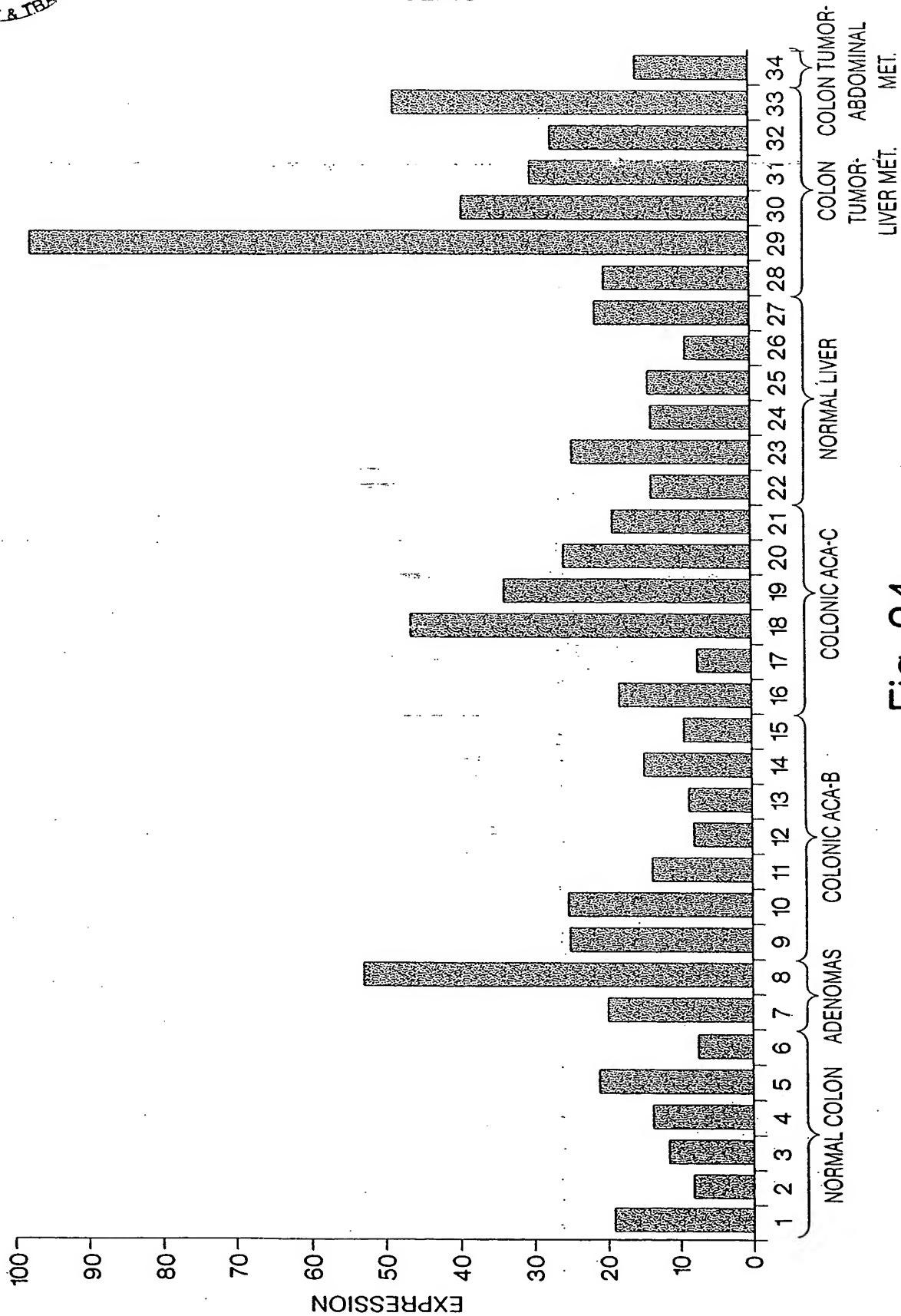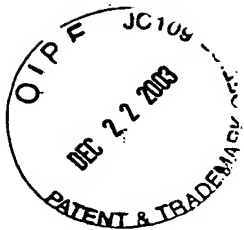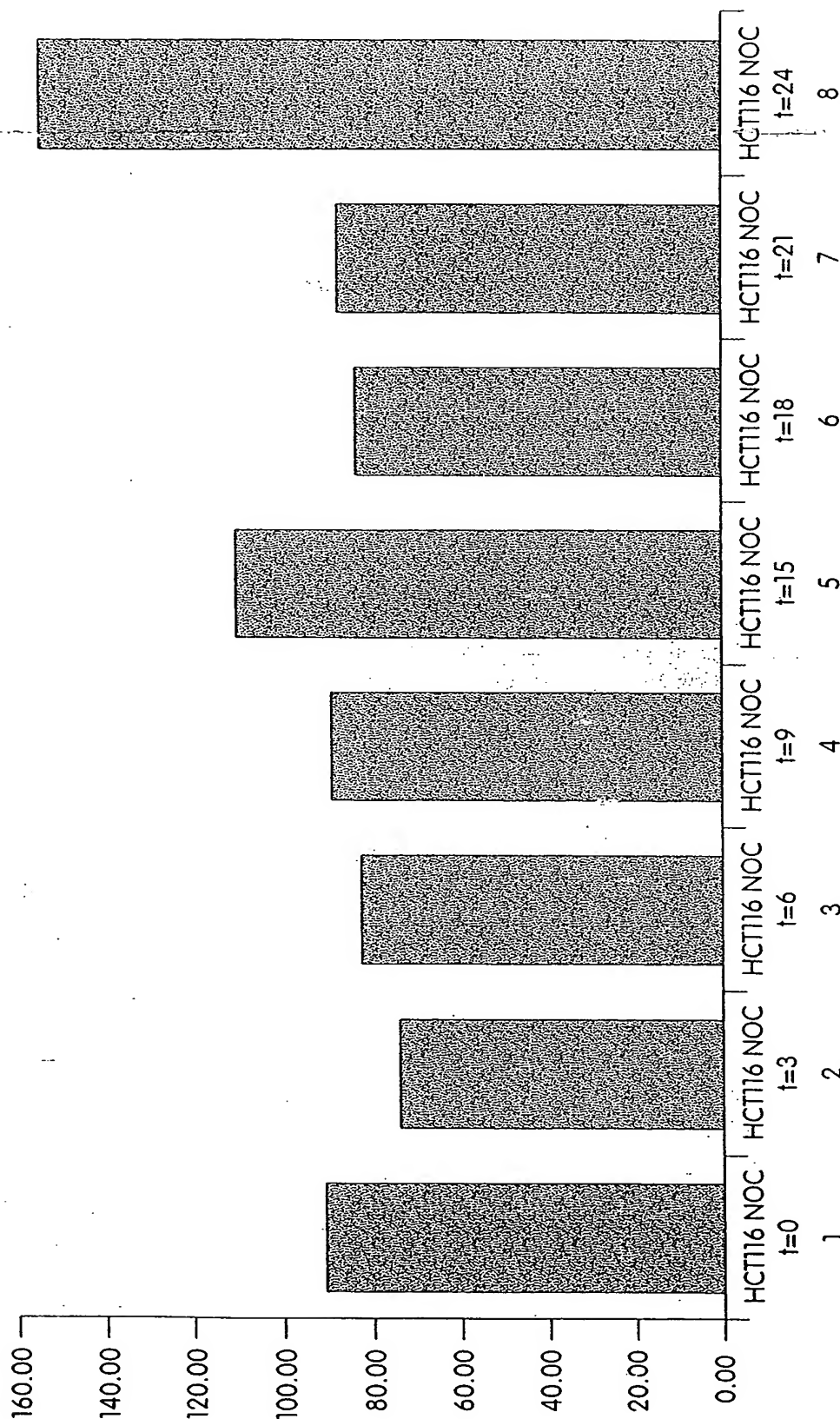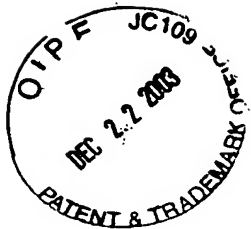

Fig. 21B

Fig. 22

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
        THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
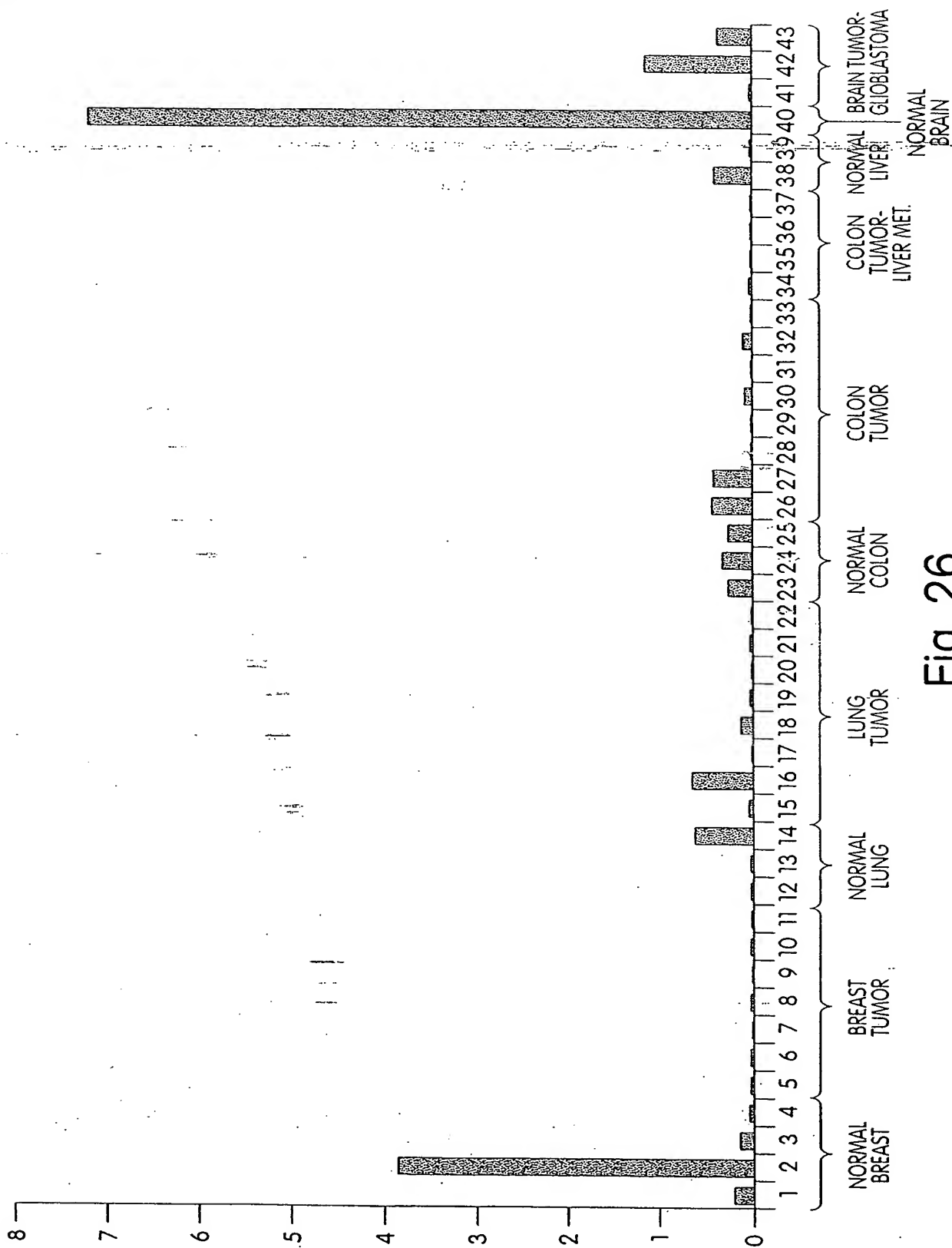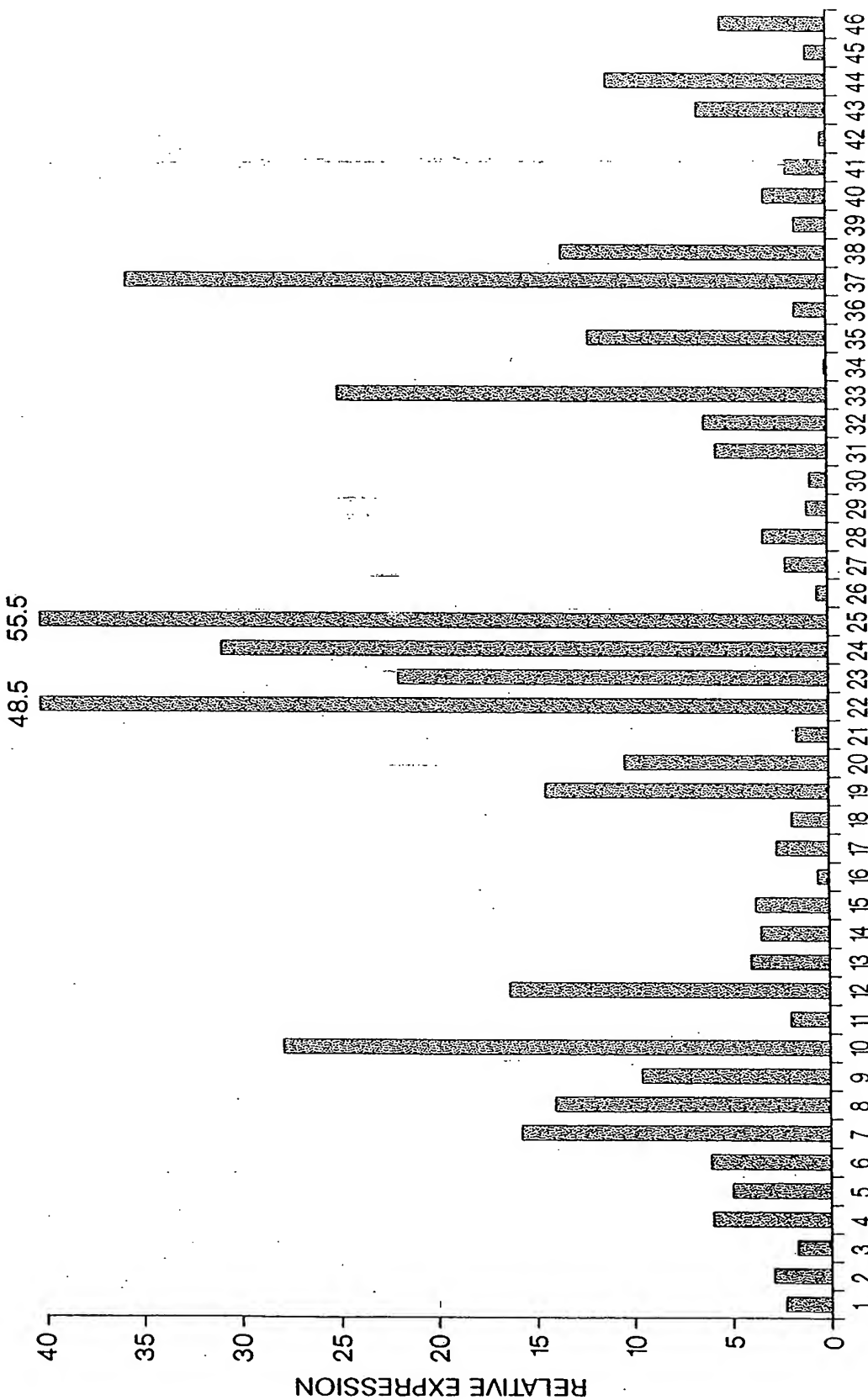   Sheet 31 of 43
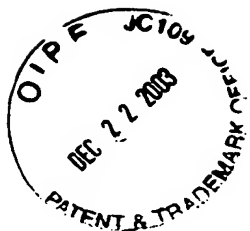   31/43

Fig. 23

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
    THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
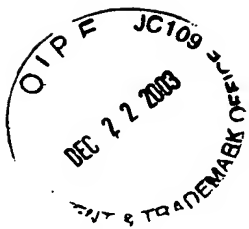Sheet 32 of 43
32/43

Fig. 24

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
        THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 33 of 43

33/43

Fig. 25

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
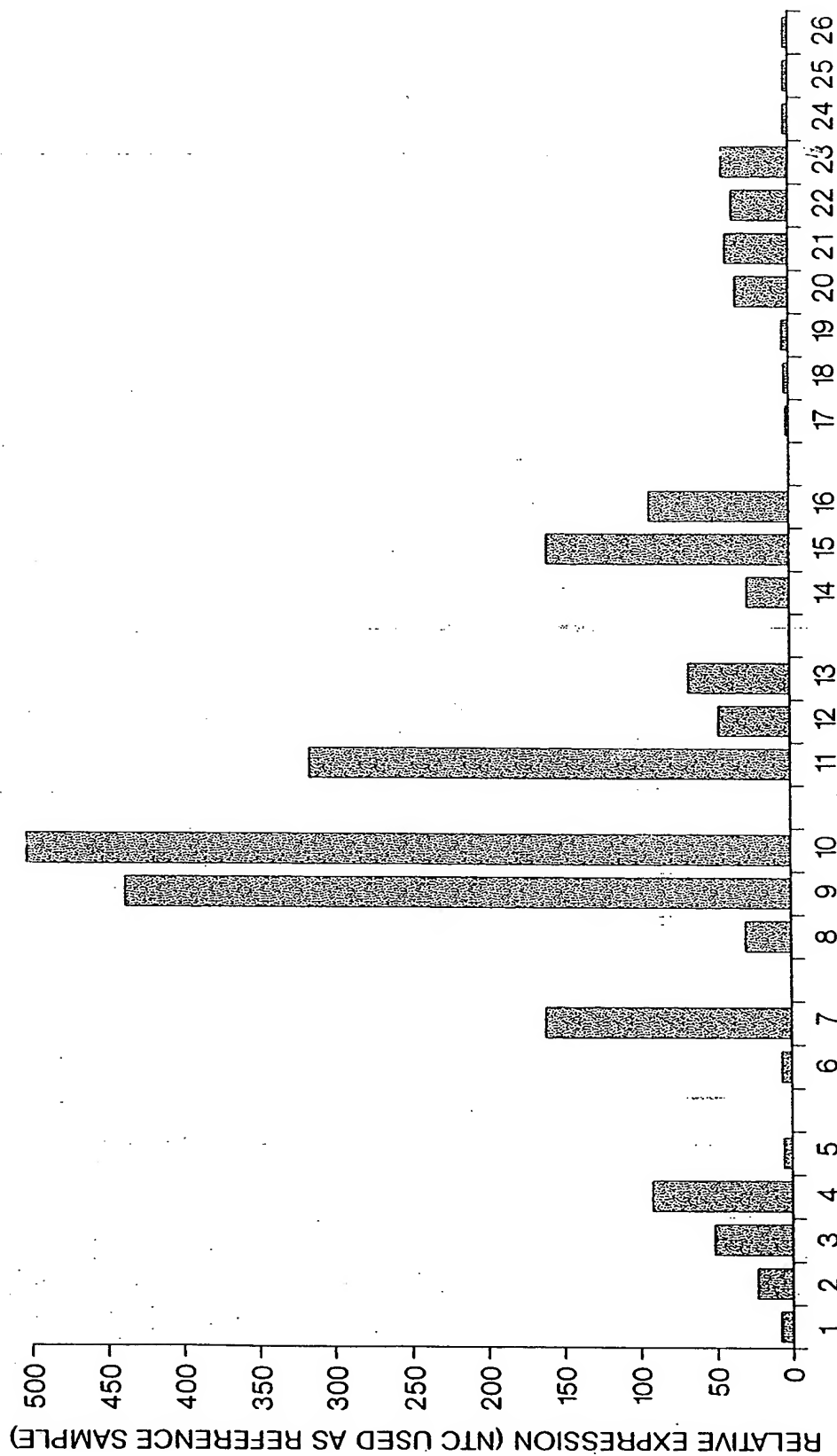Sheet 34 of 43

34/43

Fig. 26

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MPI00-079P1RCP2CN1M
Sheet 35 of 43

35/43
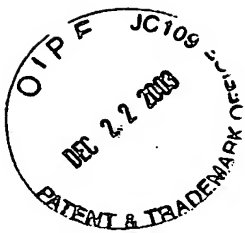
Fig. 27

Fig. 28

Fig. 29

Fig. 30

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
       THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 39 of 43

39/43

GGAATGGATGCTGTTGGCTTAAACCTCCCCCTGCCCTGGGGGTTGCAACCAGGGTCTCTG

CAAAGCCAATCCTTTGTCATCCCGCTGTCCTGCAGAGCAAGATGGGGCTCATGGCTGTCC

TGATGCTACCCCTGCTGCTGCTGGGAATCAGCGGCCTCCTCTTCATTTACCAGGAGGCAT

CCAGGCTGTGGTCGAAGTCTGCCGTGCAGAACAAAGTGGTGGTCATCACAGATGCCATCT

CAGGACTGGGAAAGGAGTGTGCTCGGGTGTTCCATGCAGGTGGGGCAAGGCTGGTGCTGT

GTGGAAAGAACTGGGAGGGACTGGAGAGCCTCTATGCCACCTTGACCAGTGTGGCTGACC

CCAGCAAGACATTCACCCCCAAGCTGGTCCTCCTGGATCTCTCAGACATTAGCTGTGTTC

AAGATGTGGCCAAAGAGGTCCTGGACTGCTACGGCTGTGTGGACATCCTCATCAACAATG

CCAGCGTGAAAGTGAAGGGGCCTGCCCACAAGATTTCCCTGGAGCTTGACAAAAAGATCA

TGGATGCCAACTACTTCGGACCCATCACTTTAACCAAAGTTCTGCTTCCCAACATGATCT

CCAGGAGAACAGGCCAGATTGTGTTAGTGAACAACATCCAAGCGAAGTTTGGAATCCCGT

TCCGCACAGCTTATGCAGCCTCTAAGCATGCCGTCATGGGCTTCTTTGACTGCCTCCGAG

CCGAGGTTGAGGAATACGATGTTGTGGTCAGCACCGTGAGCCCAACTTTCATCCGCTCCT

ACCGTGCTTCCCCTGAGCAAAGAAACTGGGAGACATCCATTTGTAAATTCTTCTGCAGGA

AGCTAGCCTATGGCGTGCACCCGGTGGAGGTGGCTGAGGAAGTGATGCGCACAGTACGGA

GGAAGAAGCAAGAGGTGTTCATGGCCAACCCGGTTCCTAAGGCTGCCGTGTTCATCCGCA

CCTTCTTCCCTGAGTTCTTCTTCGCTGTGGTGGCCTGTGGGGTGAAGGAGAAGCTCAATG

TCCCAGAAGAGGGTTAACCTCGTGGCCAAAGGGGTCACTCAAGGGGAATAAAGGCTTTCC

TAGAGAAAAAAAAAAAAAAAAAAAAAAAA

Fig. 31A
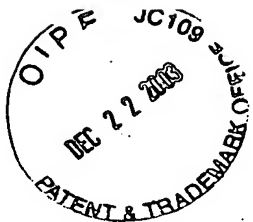
Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
THEREFOR
Attorney/Agent: Kerri Pollard Schray
Docket No.: MP100-079P1RCP2CN1M
Sheet 40 of 43

40/43

MGLMAVLMLPLLLLGISGLLFIYQEASRLWSKSAVQNKVVVITDAISGLGKECARVFHAG

GARLVLCGKNWEGLESLYATLTSVADPSKTFTPKLVLLDLSDISCVQDVAKEVLDCYGCV

DILINNASVKVKGPAHKISLELDKKIMDANYFGPITLTKVLLPNMISRRTGQIVLVNNIQ

AKFGIPFRTAYAASKHAVMGFFDCLRAEVEEYDVVVSTVSPTFIRSYRASPEQRNWETSI

CKFFCRKLAYGVHPVEVAEEVMRTVRRKKQEVFMANPVPKAAVFIRTFFPEFFFAVVACG

VKEKLNVPEEG.

Fig. 31B

Applicants: Rachel E. Meyers, et al.
Title: 21481, A NOVEL DEHYDROGENASE MOLECULE AND USES
        THEREFOR
Attorney/Agent: Kerri Pollard Schray
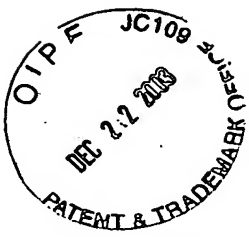Docket No.: MP100-079P1RCP2CN1M
Sheet 41 of 43

41/43

GAP of: FrGcgManager_31_UFAHDJyG_   check: 516   from: 1 to: 936

M21481 ORF - Import - vector trimmed

to: FrGcgManager_31_VFA0zr_19   check: 2871   from: 1 to: 933

h21481 ORF - Import - vector trimmed

Symbol comparison table: /ddm_local/gcg/gcg_9.1/gcgcore/data/rundata/
nwsgapdna.cmp
CompCheck: 8760

```
        Gap Weight:      12     Average Match: 10.000
     Length Weight:       4     Average Mismatch:  0.000

          Quality:    8220         Length:     936
            Ratio:   8.810           Gaps:       0
Percent Similarity: 88.103   Percent Identity: 88.103
```

Match display thresholds for the alignment(s):
               | = IDENTITY
               : =   5
               . =   1

FrGcgManager_31_UFAHDJyG_ x FrGcgManager_31_VFA0zr_19   ..

```
  1 ATGGGGCTCATGGCTGTCCTGATGCTACCCCTGCTGCTGCTGGGAATCAG 50
    |||||  |||||||| | ||||||||| ||||||||||||||||||||||
  1 atgggagtcatggccatgctgatgctccccctgctgctgctgggaatcag 50

 51 CGGCCTCCTCTTCATTTACCAGGAGGCATCCAGGCTGTGGTCGAAGTCTG 100
    |||||||||||||||||||||| |||| ||||||||||||| |||||| |
 51 cggcctcctcttcatttaccaagaggtgtccaggctgtggtcaaagtcag 100

101 CCGTGCAGAACAAAGTGGTGGTCATCACAGATGCCATCTCAGGACTGGGA 150
    | ||||||||||||||||||||||| ||||| ||||||||||||||||||
101 ctgtgcagaacaaagtggtggtgatcaccgatgccatctcaggactgggc 150

151 AAGGAGTGTGCTCGGGTGTTCCATGCAGGTGGGGCAAGGCTGGTGCTGTG 200
    ||||||||||||||||||||||| ||||||||||||||||||||||||||
151 aaggagtgtgctcgggtgttccacacaggtggggcaaggctggtgctgtg 200

201 TGGAAAGAACTGGGAGGGACTGGAGAGCCTCTATGCCACCTTGACCAGTG 250
    |||||||||||||||| | || |||| ||| |||| ||||||| ||| |
201 tggaaagaactgggagaggctagagaacctatatgatgccttgatcagcg 250

251 TGGCTGACCCCAGCAAGACATTCACCCCCAAGCTGGTCCTCCTGGATCTC 300
    ||||||||||||||||||||||||||||| ||||||||||| |||| |||
251 tggctgaccccagcaagacattcaccccaaagctggtcctgttggacctc 300
```
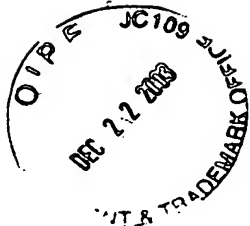
Fig. 32A

```
301 TCAGACATTAGCTGTGTTCAAGATGTGGCCAAAGAGGTCCTGGACTGCTA 350
    ||||||||| ||||||||| | ||||||||||| |||||| |||||||| |||||
301 tcagacatcagctgtgtcccagatgtggcaaaagaagtcctggattgcta 350

351 CGGCTGTGTGGACATCCTCATCAACAATGCCAGCGTGAAAGTGAAGGGGC 400
    |||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||
351 tggctgtgtggacatcctcatcaacaatgccagtgtgaaggtgaaggggc 400

401 CTGCCCACAAGATTTCCCTGGAGCTTGACAAAAAGATCATGGATGCCAAC 450
    ||||||| |||||||| |||||||| |||||||||||||||||||||||
401 ctgcccataagatttctctggagctcgacaaaaagatcatggatgccaat 450

451 TACTTCGGACCCATCACTTTAACCAAAGTTCTGCTTCCCAACATGATCTC 500
    ||||| || ||||||||| || || |||| ||||||||||||||||||||||||||
451 tactttggccccatcacattgacgaaagccctgcttcccaacatgatctc 500

501 CAGGAGAACAGGCCAGATTGTGTTAGTGAACAACATCCAAGCGAAGTTTG 550
    | ||||||||||||||||| || ||||||||| || ||||||| ||||||||
501 ccggagaacaggccaaatcgtgttagtgaataatatccaagggaagtttg 550

551 GAATCCCGTTCCGCACAGCTTATGCAGCCTCTAAGCATGCCGTCATGGGC 600
    ||||||||||||||| || |||| || |||||| |||||| || | | ||||||
551 gaatcccgttccgtacgacttacgctgcctccaagcacgcagccctgggc 600

601 TTCTTTGACTGCCTCCGAGCCGAGGTTGAGGAATACGATGTTGTGGTCAG 650
    ||||||||||||||||||||||||||||||||| || ||||||||||||| ||||
601 ttctttgactgcctccgagccgaagtggaggaatacgatgttgtcatcag 650

651 CACCGTGAGCCCAACTTTCATCCGCTCCTACCGTGCTTCCCCTGAGCAAA 700
    |||||||||||| |||||||||||| || |||| ...| |_||_|| ||||||
651 caccgtgagcccgactttcatccggtcgtaccacgtgtatccagagcaag 700

701 GAAACTGGGAGACATCCATTTGTAAATTCTTCTGCAGGAAGCTAGCCTAT 750
    ||||||||| | ||||||||| ||||||||| | ||||||||| ||||
701 gaaactgggaagcttccatttggaaattcttttcaggaagctgacctac 750

751 GGCGTGCACCCGGTGGAGGTGGCTGAGGAAGTGATGCGCACAGTACGGAG 800
    |||||||||| || ||||||||| ||||| ||||||||||| || ||||||
751 ggcgtgcacccagtagaggtggcggaggaggtgatgcgcaccgtgcggag 800

801 GAAGAAGCAAGAGGTGTTCATGGCCAACCCGGTTCCTAAGGCTGCCGTGT 850
    ||||||||||||||||||| ||||||||||| | || ||||| |||||||
801 gaagaagcaagaggtgtttatggccaaccccatccccaaggccgccgtgt 850

851 TCATCCGCACCTTCTTCCCTGAGTTCTTCTTCGCTGTGGTGGCCTGTGGG 900
    | |||||||||||||||||||||| ||||||||| |||||| |||||||||||||
851 acgtccgcaccttcttcccggagttctttttcgccgtggtggcctgtggg 900

901 GTGAAGGAGAAGCTCAATGTCCCAGAAGAGGGTTAA 936
    |||||||||||||||||||||||||||| || ||||||
901 gtgaaggagaagctcaatgtcccggaggagggg.... 933
```

Fig. 32B

GAP of: FrGcgManager_32_ZFA004eiD   check: 657   from: 1   to: 311

m21481 aa - Import - complete

to: FrGcgManager_32_AGAjaPna_   check: 9949   from: 1   to: 311

h21481 aa - Import - complete

Symbol comparison table: /prod/ddm/seqanal/BLAST/matrix/aa/BLOSUM62
CompCheck: 1102
 Matrix made by matblas from blosum62.iij

```
         Gap Weight:       12      Average Match:    2.778
      Length Weight:        4      Average Mismatch: -2.248

            Quality:     1467              Length:      311
              Ratio:    4.717                Gaps:        0
 Percent Similarity:   92.926    Percent Identity:   91.318

         Match display thresholds for the alignment(s):
                      |  =  IDENTITY
                      :  =     2
                      .  =     1
```

FrGcgManager_32_ZFA004eiD x FrGcgManager_32_AGAjaPna_ ...

```
  1 MGLMAVLMLPLLLLGISGLLFIYQEASRLWSKSAVQNKVVVITDAISGLG 50
    || ||.||||||||||||||||||| ||||||||||||||||||||||||
  1 MGVMAMLMLPLLLLGISGLLFIYQEVSRLWSKSAVQNKVVVITDAISGLG 50

 51 KECARVFHAGGARLVLCGKNWEGLESLYATLTSVADPSKTFTPKLVLLDL 100
    |||||||| |||||||||||||   ||.|| | |||||||||||||||||
 51 KECARVFHTGGARLVLCGKNWERLENLYDALISVADPSKTFTPKLVLLDL 100

101 SDISCVQDVAKEVLDCYGCVDILINNASVKVKGPAHKISLELDKKIMDAN 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 SDISCVPDVAKEVLDCYGCVDILINNASVKVKGPAHKISLELDKKIMDAN 150

151 YFGPITLTKVLLPNMISRRTGQIVLVNNIQAKFGIPFRTAYAASKHAVMG 200
    |||||||| |||||| ||||||||||||||| ||||||| ||||||| :|
151 YFGPITLTKALLPNMISRRTGQIVLVNNIQGKFGIPFRTTYAASKHAALG 200

201 FFDCLRAEVEEYDVVVSTVSPTFIRSYRASPEQRNWETSICKFFCRKLAY 250
    ||||||||||||||||:|||||||||||   ||| ||| || ||| ||| |
201 FFDCLRAEVEEYDVVISTVSPTFIRSYHVYPEQGNWEASIWKFFFRKLTY 250

251 GVHPVEVAEEVMRTVRRKKQEVFMANPVPKAAVFIRTFFPEFFFAVVACG 300
    ||||||||||||||||||||||||||:|||||::||||||||||||||||
251 GVHPVEVAEEVMRTVRRKKQEVFMANPIPKAAVYVRTFFPEFFFAVVACG 300

301 VKEKLNVPEEG 311
    |||||||||||
301 VKEKLNVPEEG 311
```

Fig. 33